Cuk- H02331-54-P014902

54

AMERICAN JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

EDITED BY

E. T. BELL
CALIFORNIA INSTITUTE OF TECHNOLOGY

ABRAHAM COHEN
THE JOHNS HOPKINS UNIVERSITY

E. W. CHITTENDEN
UNIVERSITY OF IOWA

F. D. MURNAGHAN
THE JOHNS HOPKINS UNIVERSITY

J. F. RITT
COLUMBIA UNIVERSITY

WITH THE COÖPERATION OF

FRANK MORLEY
W. A. MANNING
HARRY BATEMAN
HARRY LEVY
J. R. KLINE
MARSTON MORSE
E. P. LANE
ALONZO CHURCH
L. R. FORD
OSCAR ZARISKI
G. C. EVANS
AUREL WINTNER
GABRIEL SZEGÖ
R. L. WILDER
R. D. JAMES

PUBLISHED UNDER THE JOINT AUSPICES OF

THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

5105
.002

CALCUTTA UNIVERSITY CENTRAL LIBRARY

VOLUME LIX
1937

THE JOHNS HOPKINS PRESS
BALTIMORE, MARYLAND
U. S. A.

P14702

PRINTED IN THE UNITED STATES OF AMERICA
BY J. H. FURST COMPANY, BALTIMORE, MARYLAND

51033

INDEX.

CALCUTTA UNIVERSITY CENTRAL LIBRARY

UNIVERSITY CALCUTTA

THE COSMICAL CONSTANT AND THE RECESSION OF THE NEBULAE.[1]

By Sir ARTHUR STANLEY EDDINGTON.

In macroscopic relativity theory energy, momentum and pressure are represented by curvature of space-time; in quantum mechanics they are represented in terms of wave functions. We do not suppose that either the curvature or the waves have an objective existence. It is idle to inquire whether, when we measure energy, we are *really* measuring the bending of space-time in a fifth dimension or the frequency of oscillation of a sub-aetherial fluid. Macroscopic relativity and quantum mechanics are alternative methods of analysing our observational experience. Usually the two methods have different fields of application, and the problems treated by one method would be outside the scope of the other method. But they have a common meeting point; and I shall here consider a problem which can be solved rigorously by both methods. The problem is—to find the state of equilibrium of a radiationless self-contained system of a very large number of particles, positive and negative.

In macroscopic relativity theory this is a familiar problem which was first solved by Einstein. A self-contained static system is called an "Einstein universe." More precisely the system here treated is an Einstein universe with zero pressure; because the absence of radiation implies that the temperature is absolute zero. In quantum mechanics a radiationless steady system is said to be in its "ground state." We have therefore to determine the conditions to be satisfied by a system of N particles treated (a) as an Einstein universe with zero pressure, and (b) as a quantised system in its ground state.

The two answers must agree. But the interesting point is that the relativity solution will be expressed in terms of the gravitational constant κ and the cosmical constant λ, whereas the quantum solution will be expressed in terms of Planck's constant h and other microscopic constants. Thus a comparison of the two solutions will give us a hitherto unrecognized relation between the natural constants.

Surely the determination of this relation is the pre-eminent problem in the unification of macroscopic and microscopic theory—the unification of relativity theory and quantum theory. I shall not be able to give here sufficient

[1] A lecture given at the Harvard Tercentenary Conference of Arts and Sciences, Friday afternoon, September 4. Received by the Editors September 3, 1936.

detail for you to check the accuracy of my solution. But I will put before you the principal considerations involved; and if when you come to read the full solution[2] you are dissatisfied with the mathematics, or even (as I have heard whispered) find my treatment obscure, I cannot doubt that you will react in the normal manner of a mathematical physicist who finds an outstanding but obviously soluble problem within his reach—that is to say, you will not rest till you have solved it to your own satisfaction.

It is essential to bear in mind that the curvature and the waves are *alternatives.* We may represent a portion of the energy of a system either by curvature or by waves; but we must not include it twice over in the same formula by representing it both ways. Thus, if we use wave functions to represent part of the energy tensor, these wave functions will occupy a space whose curvature corresponds to the rest of the energy tensor. In particular, if we represent the whole energy tensor by waves, the waves will occupy a flat space. For the solution (b) of the problem enunciated we apply wave representation to the *whole* energy tensor of the system; the waves will accordingly be in flat space.

Thus our first step is to project the spherical Einstein universe into a flat space. For technical reasons it is necessary to use stereographic projection, which has the property that each element of volume remains isotropic after the projection. The uniform distribution of density in spherical space projects into a distribution concentrated towards the centre of projection and fading off to zero at infinity. This resembles the distribution of electron density in an atom. In the atom the density is concentrated towards the centre by the controlling force of the "self-consistent field"; here the controlling force is supplied by a projection factor which has the same effect mathematically. As in an atom, we must analyse the distribution into orthogonal wave functions representing possible steady states of a particle. Then, since the system is in the ground state the particles will occupy the N states of lowest energy.

It will not be necessary to investigate the individual wave functions in detail. We fix attention on two of them, namely the state of lowest energy or K state, and the highest occupied state or limit state. The K state corresponds to uniform distribution in spherical space: it is easy to calculate the wave function of the projected distribution and so determine the pressure and energy (ϵ_1) corresponding to it. The only other result that we require is a very general property of systems in a ground state, namely that when N is large the mean energy per particle is $3/5$ of the limit energy (ϵ_2), so that

[2] Eddington, ***Relativity Theory of Protons and Electrons,*** Chapter XIV.

$$\bar{\epsilon} = \tfrac{3}{5}\epsilon_2. \tag{1}$$

We already begin to see how things are going to work out. The value of N found from the recession of the spiral nebulae is of order 10^{79}. The limit state will therefore correspond to extremely high quantum number and its energy will be enormously greater than that of the K state. Actually it is found that ϵ_2/ϵ_1 is of order $\sqrt{N}$ or 10^{39}.

We might have begun the problem of unification of relativity and quantum theory by asking ourselves: What is there in common between curvature and periodicity (or frequency) which makes it possible to represent energy sometimes by curvature and sometimes by periodicity? In the simple case of a spherical distribution there is an obvious correspondence of curvature and periodicity—the periodicity being that associated with "going round the world." If there were only one particle in the universe, relativity theory would calculate its energy from the curvature (obtaining the well-known result $\tfrac{1}{2}\pi Rc^4/\kappa$), and quantum theory would calculate it approximately [3] by treating the circumference of space $2\pi R$ as the wave length of the wave function (obtaining the result $hc/2\pi R$). The two results violently disagree, the ratio being about 10^{120}; but this is not surprising, seeing that there is more than one particle in the actual universe in which h and κ are measured. Even when we divide the curvature result by $N = 10^{79}$, since it corresponds to the energy of N particles each of which should have the energy $hc/2\pi R$ due to the periodicity of its wave function, there remains a huge discrepancy. I think that this discrepancy may have frightened those who have tried tentatively to find a connection between the two ways of expressing the energy. The reconciliation lies in the fact that the periodicity $2\pi R$ corresponds to the K state of uniform distribution. Only four particles (2 positive and 2 negative) can occupy this state; and the average energy per particle is $\bar{\epsilon}/\epsilon_1$, or about 10^{39}, times larger. The exclusion effect thus completely alters the order of magnitude, and boosts the average energy up to a value which turns out to be intermediate between the energies of constitution m_pc^2, m_ec^2 of protons and electrons.

In elementary quantum theory the system that is being described—the object system—is always treated as an independent addition to the rest of the universe. If the quantum physicist ever remembers that there is a "rest of the universe," he regards it as an ideal background which will not interfere with his problem. When he writes down the ordinary wave equation for say,

[3] The exact calculation by the method of stereographic projection already described yields a result of the same order of magnitude.

two electrons, what does he suppose the other electrons in the universe are doing? Does he suppose them non-existent? If so, it will be difficult for the practical physicist to verify the equation, since he must first destroy the whole universe except two particles. It is clear that the condition that the elementary wave equation shall describe the actual performance of two electrons is, not that the rest of the universe shall be annihilated, but that it shall be arranged in some standard manner. Writers on quantum theory never tell us what standard arrangement is intended; but we can infer it from the assumptions which they commonly make. The rest of the universe must form an impermeable background with no vacant levels to swallow up unexpectedly the two object particles. Nor must any of its particles compete with the object particles for the energy levels in the added system. This means that the rest of the universe must form a system in the ground state with all the levels filled up to a certain limit energy. On top of this comes the added system—*on top,* because, the lower levels being filled, there is no room to add particles except at the top. The limit energy of the background is the threshold energy for the particles of the added system. The particles of the added system have therefore a minimum energy. This is the energy which we recognize as the rest energy or proper mass of the particles.

It must be understood that this arrangement of the rest of the matter of the universe as a system in the ground state, i. e. as an Einstein universe, is not a hypothesis about the actual state of things. (In fact we have reason to believe that the universe is now rather far removed from the Einstein state.) It is the arrangement postulated, when we apply the elementary field-free equations of quantum theory and expect the behaviour indicated by those equations to be exactly fulfilled. Of course, we do not really expect a simple equation to describe what will happen in actual conditions; in actual conditions there are all sorts of perturbing causes—additional protons and electrons, positrons, photons, electromagnetic and gravitational fields—to be taken into account. But when the rest of the universe is not causing perturbations of this kind it must still be accounted for; and it must then form a completely filled series of states below the threshold energy, providing an impermeable foundation for the added system and not otherwise interfering with it. In this way its interference with the object particles is limited to forcing their energies up to the threshold level; that is to say, it is responsible for the proper mass or inertia of the object particles. To use the terminology of relativity theory, the rest of the matter of the universe provides a *pure inertial* field. Any deviation of the rest of the universe from this standard arrangement must be explicitly described as an addition to the fixed background, and in-

cluded as such in the added system. Usually these additions are described as fields. For example, a temperature excitation of the background is described as a *field of radiation*; and exchanges of energy between the object particles and the excited background are described as exchanges between the object particles and a field of radiation incorporated in the added system. In this way we contrive formally to keep the background unchanged, and aloof from what is happening in the added system.

It will be seen that our investigation here links on to Dirac's theory of the positron, provided that certain rather obvious amendments of the latter theory are made. The absence of an electron from a sub-threshold level is represented as the *addition* of a positron to the fixed background. This is virtually Dirac's theory, except that to provide the sub-threshold levels he invented *ad hoc* an infinite quantity of matter undetectable so long as it was uniform. But it is premature to invent unrecognized matter, when there are 10^{79} particles of recognized matter pathetically appealing to the quantum physicist not to neglect them! We emphatically reject this demand for new matter. Other points on which the present theory differs from Dirac's are: (1) The number of sub-threshold levels is not infinite, and we shall presently find the precise number; (2) The levels are filled equally with protons and electrons, so that we avoid Dirac's arbitrary suppression of an infinite negative charge; (3) There is no differential treatment of protons and electrons, so that negatrons are admitted equally with positrons.

We have seen that the rest energy mc^2 of a particle in the added system is the limit energy ϵ_2. Every particle included in the object system means one particle fewer in the "rest of the universe." We can withdraw enough particles to build any system ordinarily studied without making any sensible impression on the huge number N. But cosmical problems form an exception. When we treat the whole universe as our object system, the particles cannot all come from the topmost levels of energy; and the total energy is $N\bar{\epsilon}$ not $N\epsilon_2$, i. e. $\frac{3}{5}Nmc^2$ instead of Nmc^2. The deficit $\frac{2}{5}Nmc^2$ is the amount by which the rest energy of the N particles is less than the sum of the rest energies of the particles taken singly. The ordinary name for such a deficit is gravitational potential energy. Thus $\frac{2}{5}Nmc^2$ is the negative gravitational energy of the matter forming an Einstein universe. Gravitational energy is a kind of negative exclusion effect—due to our having included the *maximum* exclusion energy in the standard mass of the proton or electron. It should be understood that we are not putting forward a new explanation of gravitation; we accepted at the outset the relativity representation of gravitational fields by curvature of space-time. But we have since united relativity and quantum

theory; and we now come across the same thing again, but viewed from the topsy-turvy outlook of elementary quantum theory.

Let us now proceed to the result of the calculation. It is comprised in the formula

$$\sqrt{\tfrac{5}{3}N} = \frac{136}{20}\,\frac{hc}{\kappa(m_p + m_e)^2}. \tag{2}$$

The source of the factor $\tfrac{5}{3}$ will be evident from what has already been said. The factor $^{136}\!/_{20}$ is introduced when we take cognizance of the charge and spin of the elementary particles. Perhaps it is simplest to describe it as an averaging factor. Two kinds of particles of unequal mass are present; and for the purposes of the present calculation the equivalent single mass is, not the arithmetic mean, but a more complicated function of the two masses which (for the particular mass-ratio concerned) is $^{20}\!/_{136}$ of the arithmetic mean.

The cosmical constant λ is connected with the total mass $\tfrac{1}{2}N(m_p + m_e)$ of the matter in the Einstein universe by the well-known formula [4]

$$\tfrac{1}{2}\pi\lambda^{-\frac{1}{2}} = \tfrac{1}{2}N(m_p + m_e)\kappa/c^2. \tag{3}$$

Thus we can, if we like, eliminate N in (2) and introduce λ in its place. We have then the promised relation between the macroscopic constants κ, λ and the microscopic constants h, m_p, m_e, obtained by solving the same problem in two ways.

Since a knowledge of N is in practice equivalent to a knowledge of λ we look on N as a deputy cosmical constant. To distinguish it, we shall call it the *cosmical number*. Its value can be found from (2) with considerable accuracy, using the observational values of the constants on the right-hand side. We obtain $N = 3{\cdot}145\,.10^{79}$. An instructive way of expressing it is

$$\tfrac{1}{2}N = 135{\cdot}82\,.2^{256}.$$

This suggested to me that the exact value of N might be $2{\cdot}136\,.2^{256}$. I have since been able to show that this is the total number of independent quadruple wave functions of the Dirac type; and an argument from epistemological considerations leaves, I think, no doubt that the principles which we follow in dissecting the sub-stratum of phenomena into elementary particles are such as to yield this total. I feel satisfied therefore that the cosmical number is precisely $2.136.2^{256}$. This number includes both electrons and protons (positrons and negatrons counting negatively in the total). I cannot say definitely

[4] This formula is given by the familiar method of solution, referred to above as method (a).

that the aggregation of some of the particles into atomic nuclei will not introduce any modification. It should however be remembered that we are primarily interested in the number of particles, not in the actual universe, but in the ideal universe postulated in those elementary equations which contain the definitions of m_p, h, etc.

Now let us turn to the extra-galactic nebulae. From (2) and (3) we obtain the cosmical constant and hence by the Friedman-Lemaître theory the limiting speed of recession of the nebulae. We have in fact made two independent determinations, since the value of N has been found in two ways. In the method which I have described at length, we employ the observed values of κ, h, c, m_p. The second method is bolder and employs no observational data, so far as N is concerned. It takes advantage of the fact that *we* lay down the conditions to be fulfilled by a fragment of that maze of interrelatedness which we call the physical universe, to qualify for the status of an independent elementary particle; and in laying down these conditions we (inadvertently) fix the total number of these particles.

From these determinations of the cosmical constant the limiting speed of recession of the spiral nebulae is—

432 kilometres per second per megaparsec.

We must allow some deduction for the gravitational attraction of the nebulae on one another, so that we may expect to observe a recession of 400 km. per sec. per mp., or perhaps rather less. The observed value is usually given as 500 km. per sec. per mp., which is in as close agreement as could be expected.

As a result of this new approach to the problem it is necessary to make two changes in the usual theory of the expanding universe. The mass M_e of an Einstein universe is given in terms of the cosmical constant by the equation $M_e = \frac{1}{2}\pi c^2/\kappa\lambda^{\frac{1}{2}}$. Confining attention to the case of zero pressure, it has been usual to distinguish three possibilities, according as the actual mass M of the universe is greater than, equal to, or less than M_e. In the present theory there is only one case $M = M_e$ to be considered. It happens that this is the choice on which I have plunged in my earlier writings, though admittedly I was guided solely by aesthetic considerations; but now the ambiguity left by relativity theory is definitely settled by quantum theory. Relativity theory gave us no reason to suppose that a collection of particles at zero temperature can form an equilibrium configuration; but quantum theory, appealing to the exclusion principle, assures us that they have an equilibrium configuration, namely the ground state. Thus universes with $M \neq M_e$, which have no

equilibrium configuration, are excluded by quantum theory. In the Friedman-Lemaître theory, it was probably not intended to deny that the universe must in any case have a ground state; but this state was pictured as a kind of "white dwarf" universe in a condition of density outside the scope of the usual cosmical equations. The present calculation shows that, on the contrary, the density of the ground state is of order 10^{-28}.

The second change is in regard to the interpretation of the cosmical constant. It was formerly supposed that λ determined a cosmical curvature $G_{\mu\nu} = \lambda g_{\mu\nu}$, which exists in entirely empty space. It is now clear that $G_{\mu\nu}$ is zero in definitely empty space (i. e. space in which the probability that a particle or photon is present is zero). Definitely empty space is a rather inconvenient abstraction; but we may express the result in a more concrete way by saying that in the space between the galaxies, where the density is much less than the average density of matter in the universe, the curvature is less than $\lambda g_{\mu\nu}$. I cannot enter fully into the reasons for this change of view; but the following hint will perhaps suffice. We had to catechize the quantum physicist, who writes down a wave equation for two or three particles, as to what he had done with the rest of the universe. Similarly when the cosmologist treats the curvature of a vacuum, we have to ask what he has done with the particles removed. In order that we may represent material bodies or vacua as formed by aggregating or dispersing particles (without creation or annihilation of particles) we describe them as modifications of an initial standard probability distribution. In current quantum theory this standard probability distribution (which must not be confused with the standard distribution as an Einstein universe referred to earlier) is known as the *a priori* probability; and the wave functions represent the modifying factor required to transform it into the actual distribution. Since the standard distribution itself is not represented by wave functions, its energy, etc. must be represented by curvature. The cosmical curvature corresponds to this standard or *a priori* probability distribution, and not to a vacuum.

CAMBRIDGE UNIVERSITY,
ENGLAND.

THE RELATIVISTIC PROBLEM OF SEVERAL BODIES.[1]

By Tullio Levi-Civita.

1. A key stone of classical mechanics is Newton's third law or law of reaction. An immediate consequence of it is that, for every body, the motion of the centre of mass goes on as if this fictitious point were a real particle, acted on all forces *external* to the body. We may therefore disregard, as far as the motion of this point only is concerned, any interior action between the material elements of the body itself.

No analogous statement holds in general relativity. You may still, like in the old mechanics, write down, for every infinitesimal element, the equations of motion corresponding to given physical circumstances. Then, having in view a set of elements, say a natural body, you may form as usual a linear combination of the individual equations, thus introducing the motion of the centre of mass; but in this combination internal forces are no longer eliminated. This indicates that, in general relativity and especially in attacking on astronomical point of view the problem of several bodies, the abstraction of material points (infinitesimal sizes, finite masses) does not offer the same convenience, nor even the same legitimacy, as usual.

Indeed we would meet essential difficulties in searching to pass to the limit too early. The natural way is to start from the general relativistic field-equations, which control the ds^2 and the motion of a *continuous* crowd of material particles.

2. Suppose that, in a given four-dimensional domain of variables, x^0 (timelike), x^1, x^2, x^3 (spacelike) coördinates, there exist a metrics and a flow (vector-field). Mathematically we have then a differential form

$$ds^2 = \sum_{0}^{3}{}_{ik} g_{ik} dx^i dx^k; \tag{1}$$

a timelike unit world-vector $\boldsymbol{\lambda}$, whose contravariant and covariant components are respectively

$$\lambda^i = dx^i/ds, \qquad \lambda_i = \sum_{0}^{3}{}_{k} g_{ik}\lambda^k \qquad (i = 0, 1, 2, 3) \tag{2}$$

[1] A lecture given at the Harvard Tercentenary Conference of Arts and Sciences, Friday forenoon, September 4, 1936. Received by the Editors September 3, 1936.

verifying—with the usual notations of the absolute differ
identities

$$\sum_0^3 {}_i \lambda_i \lambda^i = \sum_0^3 {}_{ik} g_{ik} \lambda^i \lambda^k = 1; \tag{3}$$

and besides a scalar density $\epsilon(x^0, x^1, x^2, x^3)$ fixing at any i
tion of energy and therefore also, save for a constant factc

Field and motion appear accordingly to involve 15
independent variables: the ten g, the four λ and ϵ. As
by one algebraic relation (3), we want essentially a systen
conditions.

If we suppose that only gravitation has to be taken int
cal stresses behave as if the moving matter be incoherent, an
tensor reduces then to the form

$$T_{ik} = \epsilon \lambda_i \lambda_k \tag{4}$$

With this tensor, Einstein's gravitational equations and th
sequences afford us the required conditions. Indeed, with
the ten gravitational equations

$$G_{ik} - \tfrac{1}{2} G g_{ik} = - k T_{ik} \tag{I}$$

imply the conservation principles, namely the fact that tl
of the tensor T_{ik} must vanish, which is expressed by

$$\sum_0^3 {}_k T_{ik}|^k = 0$$

Substituting for T_{ik} their values (4), we get

$$\lambda_i (d\epsilon/ds + \epsilon \operatorname{div} \boldsymbol{\lambda}) + \epsilon p_i = 0, \tag{5}$$

where

$$p_i = \sum_0^3 {}_k \lambda_{ik} \lambda^k \tag{6}$$

are the covariant components of the vector $\boldsymbol{p}$ (geodesic cu
line, i. e. trajectory, passing through the point).[2] Equ
firstly that, everywhere matter is present, that is $\epsilon \neq 0$, t
desic, in as much as their vector of curvature must vanish.

[2] Compare e. g. my *Absolute Differential Calculus* (Blackie, 1927), pp. 135 and 362.

is in any case orthogonal to $\boldsymbol{\lambda}$; therefore multiplying (5) by λ^i and summing, it results, owing to (3),

(III) $$d\epsilon/ds + \epsilon \operatorname{div} \boldsymbol{\lambda} = 0,$$

which is the *equation of continuity* for our gravitational flow. The system (5) reduces then simply to

(II) $$\epsilon p_i = 0 \qquad (i = 0, 1, 2, 3),$$

stating, for $\epsilon \neq 0$, the geodesic property of every trajectory. On the other hand, in empty space ($\epsilon = 0$) trajectories have no physical meaning; they are indeterminate, the counterpart of it being that the equations (II) are automatically satisfied for $\epsilon = 0$.

The four-dimensional equation of continuity (III) gives rise, in a very simple and general manner, to the relativistic correction of the ordinary principle of invariableness of gravitational mass throughout the motion. Indeed (III) is the mathematical expression of the fact that $\int \epsilon dV$ (dV elementary hypervolume) extended to a portion of the fourfold variety (x^0, x^1, x^2, x^3), having the metrical determination (1), is an (integral) invariant of the flow along the geodesic lines. On the other hand the ds^2 is a differential invariant of the same motion, or analytically of the differential system $p_i = 0$ defining the geodesic lines. We have accordingly, along every world-line,

(7) $$\epsilon\, dV/ds = \text{const.}$$

Now the four-dimensional extension dV may be split up in the product

$$dV = \sqrt{-g}\, dx^0 dx^1 dx^2 dx^3 = \sqrt{g/g'}\, dx^0 \sqrt{-g'}\, dx^1 dx^2 dx^3,$$

g and g' indicating respectively the determinants of ds^2 and of its spacial part $(ds^2)_{dx^0=0}$. As

$$\epsilon \sqrt{-g'}\, dx^1 dx^2 dx^3$$

is the energy, or, except for a constant factor, the ordinary gravitational mass dm carried by the moving element, and, with obvious notation,

$$g'/g = g^{00},$$

the preceding equation (7) takes the form

(8) $$\frac{1}{\sqrt{g^{00}}}\, dm \frac{dx^0}{ds} = dm_0 \qquad (dm_0 \text{ constant along the motion})$$

reducing for pseudoeuclidean space to

$$(8') \qquad dm = dm_0 \sqrt{1-\beta^2}$$

where [3] β means the velocity of the moving particle (x^0 performing the function of time); and the constant dm_0 its ordinary or rest-mass.

In conclusion the mathematical aspect of the most general relativistic problem of the motion of gravitating media consists of 14 partial differential equations between as many unknown functions. As such general media do not exclude the case of separate bodies, ϵ being 0 outside, we have at the same time mathematically sketched even the relativistic problem of several bodies.

3. This rigorous position has been and probably will be successful only in a few cases. Firstly in the celebrated Einstein-Schwarzschild one-centre problem, in which the ds^2, for the space exterior to the centre, may be inferred from (I) through symmetry requirements. Geodesic principle (reduced then to ordinary differential equations) determines afterwards the motion of an *infinitesimal* body in this gravitational field.

Secondly, in the more recent investigations concerning the expanding universe, where some exact solutions have been detected, again in the hypothesis of spherical symmetry (around our immediate surrounding).[4]

Some other exact solutions, especially for static ds^2, have been determined, but their circumstances are not realized, or at least have not hitherto been found realized, in any astronomical question.

4. In order to attack, according to relativity, the usual problems of celestial mechanics, besides the Einstein-Schwarzschild one-centre problem, only approximation's procedures have been tried, and seem to be at present available. They have been initiated by Einstein himself,[5] who integrated, through retarded potentials, the field-equations (I), retaining only linear terms. This has led, among others, to ascertain the existence of gravitational waves.

Almost at the same time the integration by successive approximations of the general equations of moving and gravitating matter has been outlined by *J. Droste* [6] in a thorough but extremely concise paper. The orders of magni-

[3] It is not out of place to remark that, for the sake of motion, gravitational mass suffers an alteration (8′), which is exactly inverse of that experienced by inertial mass.

[4] Sources of reference, emphasizing the mathematical point of view, are: G. C. McVittie, *Monthly Notices of R. A. S.*, vol. 93 (1933), pp. 325-339; J. L. Synge, *Proceedings of the National Academy of Sciences*, vol. 20 (1934), pp. 635-640; C. Tolotti, *Rendiconti dell' Accademia dei Lincei*, vol. 21 (1935), pp. 326-331, 488-492, 571-575.

[5] "Näherungsweise Integration der Feldgleichungen der Gravitation," *Sitzungsber. der Preuss. Ak. der Wiss.*, 1916, pp. 688-696.

[6] "The field of moving centres in Einstein's theory of gravitation," *Ak. van Vet. te Amsterdam*, vol. 19 (1926), pp. 447-455.

tude are here previously established with reference to usual conditions of our planetary system; and the process of calculation is shown, for the problem of several, say $n+1$ bodies, with some useful remarks intended to simplify the laborious task of obtaining, up to the first order, the relativistic corrections in the ordinary differential equations.

This task has been absolved by the late *De Sitter* in his comprehensive researches "On Einstein's theory of gravitation and its astronomical consequences."[7]

In the Newtonian theory, as mentioned at the beginning, the motion of the centre of mass P_h of the h-th body C_h is influenced by the attraction of the other bodies, *not of C_h itself.* To get the surrogate of this, even only in the approximate relativistic scheme, is not a simple straightforward matter. De Sitter's survey duly explains the spirit of Einstein's theory, and carefully prepares explicit formulae for eventual astronomical applications. One point however, the contribution of every body to the motion of its material elements, is sketched too hastily. The contributions are not ignored: on the contrary; but the admissions permitting to reduce them to a minimum (of residual constant parameters) are not patiently stated, and even less followed step by step.

Thus it happened that De Sitter was led to the conviction that the self-contributions would only require a slight modification of the constant values of masses, after which no other influences remain in the final equations.

I have recognized after accurate discussion, about which I shall say a few words later on, that the final ignoration of self-contributions is indeed realizable for the problem of the *two* bodies, but not for three or more.

As a matter of fact it may be added that, while the wide interest of De Sitter's work has been soon acknowledged, authoritative voices asked early[8] for a more cogent investigation concerning what, in Newtonian language, would be the residual effects of interior forces on the general motion of a body.

This reserve brought substantially to the diffuse belief[9] that De Sitter's formulae are certainly suitable to govern the events of an *infinitesimal* body, in the field of any others *in given motion,* but not, without caution, to predict the mutual influences of a system of celestial bodies.

---

[7] *Monthly Notices of the R. A. S.*, especially second paper, vol. 67 (December 1916), pp. 155-183.

[8] Compare especially M. Brillouin, "Gravitation Einsteinienne. Statique. Points singuliers. Le point matériel," *Comptes Rendus*, T. 175 (1922), pp. 1008-1012.

[9] This point of view is adopted e. g. in the excellent book of Prof. J. Chazy, *La théorie de la relativité et la mécanique céleste,* Paris, Gauthier-Villars, T. 2 (1930).

5. It is now time to account for the several admissions, which enable firstly the approximate integration of field-equations (I), i. e. the determination of the ds^2; and afterwards the construction of the equations of the motion of the centres of mass $P_0, P_1, \cdots, P_n$ of the bodies, with the aim of getting rid of any other unknown function of x^0 (time) and to reach finally ordinary differential equations as in the classical problem of $n+1$ bodies.

The adopted criteria of simplification are of four kinds and may be summed up as follows:

A_1) Neglect of quantities of order higher than one, in a well defined sense, suggested by the circumstances of our planetary system. Accordingly we shall have to do with velocities of material bodies small in comparison to the velocity of light: order 10^{-4}, as for the earth. For astronomical purposes it will be obviously sufficient to secure the relativistic corrections up to 10^{-8}, and we shall assume this as our need of accuracy: attributing the *first order* to terms having such a small rate to some other standard term in the same equation or relation. We shall generally put, for the ordinary velocity v,

$$v/c = \beta,$$

and say accordingly that β has the order 1/2. In as much the adopted variables are: x^0, which means (nearly) ct (t ordinary time); and x^i $(i > 0)$, which represent (nearly) cartesian coördinates, we may consider for every material particle: dx^i/dx^0 a dimensionless quantity of order 1/2, or $\sim \beta$, $\sim$ meaning "of the same order as . . ."

Furthermore, $f(x^0, x^1, x^2, x^3)$ being a function which varies with x^0 only because of motion of material bodies, $\partial f/\partial x^0$ is of order 1/2 higher than $\partial f/\partial x^i$ $(i > 0)$.

Last but not least comes the behaviour of

$$ds^2 = \sum_0^3{}_{ik} g_{ik} dx^i dx^k.$$

In empty space the coefficients reduce to the $g^0{}_{ik}$ of the Einstein-Minkowski form $dx^{0^2} - (dx^{1^2} + dx^{2^2} + dx^{3^2})$.

Putting

$$g_{ik} = g^0{}_{ik} - 2\gamma_{ik} \qquad (i, k = 0, 1, 2, 3), \tag{9}$$

all γ_{ik} are to be treated as small quantities of the first order at least; γ_{0i} $(i > 0)$ even of order not beneath 3/2, this arising from the precision with which could be ascertained the isotropy of light propagation through sky, notwithstanding the presence of celestial bodies.

An important complement of these qualitative preliminaries is to avoid superfluous calculations. Aiming to assure the correctness up to the terms of order one in the final equations of motion, it is required, as first remarked by Einstein himself and easy to verify, to procure only the lowest terms of the γ_{ik}, excepting for γ_{00} which is required up to the second order inclusive.

6. Even in the classical mechanics, the problem of several bodies may be reduced to ordinary differential equations only using, besides exact consequences of the mechanical laws, the circumstance (generally, if not always verified) that the dimensions of the bodies are very small in comparison with the mutual distances. The precise assumption, allowing an autonomous treatment of the motion of the centres of mass, is that, D being the largest dimension of any one of the given bodies, and R a lowest bound for mutual distances between points of different bodies, during the motion, not necessarily D/R, but

A_2). $(D/R)^2$ *is completely negligible.*

As relativity includes ordinary mechanics as limiting case, the same condition A_2) will certainly be required in order to disentangle the motion of the centres of mass. And we shall accordingly associate A_2) to A_1).

There is however an important feature of the ordinary conception, which is lost in applying A_2) to general relativity. In Newtonian mechanics nothing prevents us to suppose the dimension D of the bodies to become smaller and smaller, their mass remaining finite. On the contrary the relativistic determination of the g_{ik} in astronomical problems needs, just in the first approximation, the introduction, for every body C_h, of its own potential, precisely in some point P_h inside C_h, namely

$$\int_{C_h} \frac{\mu' dC'}{r(P', P_h)},$$

where μ' represents the density of the body in its point P', and $r(P', P_h)$ the distance between P' and P_h. Such a value ($> m_h/D$, m_h being the mass of C_h) tends clearly to infinity by lessening the dimensions of C_h and maintaining finite its mass m_h. It is therefore a wise policy to avoid such difficulties, placing ourselves in the circumstances of celestial bodies, for which $(D/R)^2$ may be safely neglected in seeking after first order effects of relativity.

On the other hand it is well to bear in mind that a still greater approximation could not be attainable (as is, on the contrary, the case in ordinary mechanics) by considering bodies of infinitesimal sizes and finite masses: we are then exposed, as just mentioned, to the incongruence of something becoming infinite.

7. The admissions A_1) and A_2) are undoubtedly essei to perform calculations; they lead especially to the kno proposed accuracy, of the ds^2, whose coefficients result as dep velocities and accelerations of all moving elements. It is plying geodesic principle, to write down, for any one of equations of motion. Such equations of course do not ordinary differential system. Truly they concern generic p media (our $n+1$ bodies) and belong to the much more c differential domain.

The further reduction to ordinary equations resemblin mechanics is spontaneously suggested by the last, where, motion of centres of mass P_h is rigorously controlled by a differential equations.

The obvious suggestion is to combine the functional eq at, which define the acceleration of any material element, the acceleration of the various centres of mass P_h ($h=0$, will consist of Newtonian terms + relativistic corrections. corrections, even in the reduced form which follows from A_1 upon the motion, not only of the P_h's, but in general also o bodies C_h. To remove such a priori unknown influences, must be added. Fortunately, as the disturbing bodily infl in the relativistic corrections (order 10^{-8}), it may be ant suppositions (verified within one or a few hundredths) w the required first-order accuracy, as far as 10^{-2} of a first o a negligible amount.

The hypothesis, in virtue of which one succeeds in eli tional residua, are two: one, A_3), of kinematical and structural contents.

8. A_3) We shall admit that *the motion of every l roughly a pure translation.* Roughly means here that, if sents the velocity of P_h and $\beta_h + \Delta\beta_h$ the velocity of any same body C_h, the oscillations are not relevant; more preci absolute values

$$|\Delta\beta_h| / |\beta_h|$$

never goes beyond a few hundredths. It is certainly so for our Earth especially the speed due to rotation at the 1/60th of the translational one; not to speak of eventual timerates are in comparison absolutely negligible.

An obvious consequence of A_3') is that, in the same order of approximation, the bodies behave as rigid ones; therefore not only configurations and densities are preserved, but moreover the Newtonian potential of any body C_h will have, in interior points P_h of the same body, invariable values. Accordingly we shall be allowed to treat (in the relativistic corrections) *as independent of* x^0 expressions like

$$(10) \qquad \varpi_h = \frac{f}{c^2} \int_{C_h} \frac{\mu' dC'}{r(P', P_h)},$$

where f is the constant of gravitation, μ', as before, the material density.

Except for a constant homogeneity factor (converting the integral into a pure number), ϖ_h is obviously the value of the potential of the h-th body in its centre of mass P_h.

Similarly every C_h will bear almost (always within a few hundredth) unchanged, during the motion, the energy of its own Newtonian distribution, as well—with a far greater accuracy assured by (8′)—its mass m_h. Therefore we are enabled to treat as true constants the dimensionless first order quantities

$$(11) \qquad \eta_h = \frac{1}{m_h} \frac{f}{c^2} \int_{C_h} \mu(P)\, dC_P \int_{C_h} \frac{\mu(P')}{r(P', P)}\, dC_{P'},$$

9. The ϖ_h's, just spoken of, though constant with respect to the time x^0, are functions of the interior point P_h, and will thus admit in this point partial derivatives, in general not all zero. Accordingly, in the Lagrangian function belonging to the point P_h, ϖ_h cannot be treated as constant, for, in constructing the equations of motion, the ϖ_h's must be derived just with respect to the coördinates of P_h. Such a complication disappears as soon as the vector $\mathrm{grad}_{P_h} \varpi_h$, proportional to Newtonian force, vanishes in P_h. This is obviously the case if the body admits geometrical and material symmetry around one of its points, which is then of course the centre of mass P_h.

But it is not necessary that $\mathrm{grad}_{P_h} \varpi_h$ is rigorously zero. It suffices for us that this vector might be neglected in the valuation of relativistic correction. As an essential part of it arises, for every P_h, from the external force due to the other bodies besides C_h, the practical conclusion that in our first order approximation ϖ_h may be treated as a true constant (independent as well of x^0, as of the position of P_h) requires only that:

A_4) *The Newtonian attraction of the body C_h on its centre of mass P_h amounts at most to some hundredth of the attraction exerted, on the same point P_h, by all other bodies of the system.* This will be our fourth and last admission. By the cumulative aid of the four, we are at last enabled to

recognize that the motion of the P_h's finds its reduced mathematical translation in a system of ordinary differential equations of the same total order as the Newtonian problem of the $n+1$ bodies. The ϖ's and η's are to be treated as intrinsic constants of the bodies, well defined in every concrete case of stars or planets. They possess however, as already remarked, the singular character of tending to infinity with the concentration of matter, namely when size decreases, mass remaining finite. This behaviour has no analogue in the ordinary mechanics, but had been noticed in pre-maxwellian and maxwellian electromagnetism and was only recently removed by the new Born's theory.

10. It is not possible, nor would it be suitable to entertain the present audience with the rather long developments leading to an explicit form of the Lagrangian function L_h which furnishes in the usual way the three equations of the motion of the point P_h. I must confine myself to report in synthetical way the final result, to be interpreted as if the surrounding astronomical space be rigorously euclidean; the space-coördinates x^1, x^2, x^3, be cartesian with reference to a whatever galilean frame; the time-coördinate x^0 means ct, as previously stated.

Of course a choice of variables of this kind must be rooted in the very nature of the space-time metrics. I beg you to give me credit that it is really so.

To approach the end I introduce the Newtonian potential, divided by c^2, of the n bodies other than C_h, putting

$$\gamma_h = \frac{f}{c^2} \sum_0^n{}'_\nu \frac{m_\nu}{r(P_h, P_\nu)}, \tag{12}$$

where m_ν denotes the ordinary rest-mass of the body C_ν, and Σ' means that in the sum the term corresponding to the value h of the index ν must be omitted. The coördinates of the point P_h being x_h^i $(i=1,2,3)$ and $\dot{x}_h^i$ or $\beta_{h/i}$ their derivatives dx_h^i/dx^0 with respect to x^0, I shall denote by $\boldsymbol{\beta}_h$ the vector-velocity, by

$$\beta_h^2 = \sum_1^3{}_i \dot{x}_h^{i2} \tag{13}$$

the square velocity of P_h, and by

$$N_h = \tfrac{1}{2}\beta_h^2 + \gamma_h$$

the Lagrangian function of its (absolute) motion in the Newtonian problem of $n+1$ bodies. This function is exactly the classical one, excepted for the

substitution of $x^0 = ct$ to t, which is nothing but a change of unit. It implies division by c^2 of the Newtonian function of force, as shown by the expression (12) of γ_h, which has thus automatically acquired the advantage of being (like β_h) dimensionless.

Coming now to the relativistic correction to N_h, it will be convenient to split it up into two parts D_h and S_h. The first goes up to *Droste* and *De Sitter*[10] (and for that reason I use the letter D) and results from relativistic influences *except self-contributions,* of which the two authors had got rid of with excessive ease.

Professor *M. Brillouin*[11] had happily proposed to name *effacing-principle* (principe d'effacement) the systematic ignoration of self-contribution, ignoration which holds rigorously in Newtonian mechanics, because of equality between action and reaction. In default of this, self-influences must be duly calculated. What, within our approximation, comes out in this way forms the term S_h. The explicit expression of D_h is

$$
\begin{aligned}
D_h = & -\gamma_h{}^2 + \tfrac{1}{2}N_h{}^2 + \gamma_h\beta_h{}^2 \\
(15)\qquad & + \frac{f}{c^2}\sum_0^n{}'_\nu \frac{m_\nu}{r(P_h, P_\nu)}\left(-\gamma_\nu + 3/2\beta^2{}_\nu - 4\boldsymbol{\beta}_h \times \boldsymbol{\beta}_\nu\right) \\
& + \frac{f}{2c^2}\ \frac{\partial'^2}{\partial' x^{02}}\sum_0^n{}'_\nu\ m_\nu\, r(P_h, P_\nu),
\end{aligned}
$$

where $\times$ means scalar product and the dash in the operator $\partial'^2/\partial' x^{02}$ alludes to the circumstance that the derivation does not affect the point P_h, concerning on the contrary all others P_ν $(\nu \neq h)$.

The self-contribution S_h, omitting an additional constant which does not affect the Lagrangian equations of motion, has the form

$$
\begin{aligned}
S_h & = -3/2\varpi_h\beta_h{}^2 - \frac{f}{c^2}\sum_0^n{}'_\nu \frac{m_\nu}{r(P_h, P_\nu)}(\eta_\nu + 2\varpi_h) \\
(16)\qquad & = -3/2\varpi_h\beta_h{}^2 - \frac{f}{c^2}\sum_0^n{}'_\nu \frac{m_\nu}{r(P_h, P_\nu)}\eta_\nu - 2\varpi_h\gamma_h,
\end{aligned}
$$

η_h and ϖ_h being certain gravitational parameters of the h-th body, which (as already emphasized) play the rôle of (small) constants. By comparing S_h with N_h we see that S_h is built up by exactly the same inertial and gravitational terms as N_h, each affected by a small coefficient ϖ or η. The presence

[10] To be remarked however that *De Sitter* had in view our planetary system, and had accordingly abridged the final expressions of the components of the perturbative forces, treating the masses m_ν $(\nu > 0)$ as infinitesimal in comparison to m_0.

[11] *Loc. cit. ante*, p. 5.

of S_h in L_h has thus barely the effect of altering the inertial and gravitational masses of the bodies. The equality for each body of these two masses is inherent to the Newtonian term N_h, but does no more necessarily hold for the self-terms S_h. Therefore the S_h's will in general produce genuine perturbations of exactly the same order as those already signalled by *De Sitter* and synthetically collected in our D_h.

There is however an important case, the simplest one of the two bodies $(n = 1)$, in which a slight alteration of the two masses m_0 and m_1 suffices to reestablish the effacing-principle through elimination of the S_h's. To see this clearly, let me begin by the remark that a Lagrangian function whatever and particularly our

$$L_h = N_h + D_h + S_h$$

may be multiplied by any constant factor without altering the variational equations (equations of motion) arising from it.

We shall employ a factor of the form $1 + \epsilon_h$, ϵ_h being a small constant of the first order. Such a multiplication does not alter, up to the second order, quantities already of second order, like D_h or S_h. We may accordingly assume as Lagrangian function of the motion of P_h instead of L_h,

$$L^*_h = N_h + \epsilon N_h + D_h + S_h$$

with the purpose of opportunely disposing of the first order constants ϵ_h $(h = 0, 1, \cdots, n)$.

In the case of two bodies the sums in the expressions (12) and (16) of γ_h and S_h reduce to a single term, bearing, we may say, the index $h + 1$, with the obvious agreement of identifying indices differing by 2 (or a multiple of 2). If furthermore we denote the mutual distance $r(P_h, P_{h+1})$ simply by r, we may then write

$$N_h = \tfrac{1}{2}\beta_h^2 + \frac{f}{c^2}\,\frac{m_{h+1}}{r},$$

$$S_h = -\varpi_h \tfrac{3}{2}\beta_h^2 - \frac{f}{c^2}\,\frac{m_{h+1}}{r}(\eta_{h+1} + 2\varpi_h) \qquad (h = 0, 1)$$

from which, adopting $\epsilon_h = 3\varpi_h$ and putting

$$(17) \qquad m^*_{h+1} = m_{h+1}(1 - \eta_{h+1} + \varpi_h),$$

which is equivalent to

$$(17') \qquad m^*_h = m_h(1 - \eta_h + \varpi_{h+1}),$$

we get

$$L_h = (1 + 3\varpi_h)N_h + D_h + S_h = \tfrac{1}{2}\beta_h^2 + \frac{f}{c^2}\,\frac{m^*_{h+1}}{r} + D_h.$$

In the third term D_h would still appear the old rest-masses m_0, m_1, but it is allowed to replace them by the modified masses m^*_h, since the differences $m^*_h - m_h$ are of the first order and D_h already of the second.

Suppressing asterisks, now useless, because only modified masses subsists, we have finally, for the absolute motion of centres of mass in the problem of the two bodies the Lagrangian functions

$$L_h = N_h + D_h \qquad (h = 0, 1), \tag{18}$$

where $N_h = \frac{1}{2}\beta_h{}^2 + \gamma_h$ and D_h has the expression (15).

On the contrary, in the problem of three or more bodies, the self-term S_h may not be removed by the simple expedient of such slight mass modifications. At any rate we do not forget the overwhelming circumstance that, even in classical mechanics, the general solution of the problem of several bodies is not known as soon as the bodies are three or more, whilst the two bodies problem had been integrated since Newton himself.

Therefore, though looking only at first order relativistic corrections, the situation is quite different in the general case and in the two bodies problem. For the former only partial investigations may be devised with the aid of some further exact or approximate hypothesis, and so we have meanwhile not attempted to go beyond the precise establishment of differential equations.

For $n = 1$ (two bodies) it follows from the elements of celestial mechanics that only quadratures are needed to obtain first order corrections for the troubled motion, defined by the Lagrangian function (18). It is simply a matter of technical skill to deduce the final formulae in not too tedious a way, thus preparing and discussing possible astronomical tests. I shall briefly sketch and illustrate numerical results on a next occasion.

UNIVERSITY OF ROME.

. . . .

In a discussion on the general lines of this paper, which took place October 7 in Princeton, Professor Einstein expressed a preliminary doubt concerning the representation (4) of the T_{ik}. This expression corresponds to incoherent matter, thus disregarding any kind of interior action, normally not at all negligible within matter in bulk.

Professor Einstein remarked that, in order to draw consequences concerning celestial bodies, which behave like concrete aggregates and not as cosmical dust, it seems necessary to introduce in the T_{ik}'s, besides $\epsilon\lambda_i\lambda_k$, which is the relativistic translation of purely kinetic stresses, some further term

1902

concerning material stresses, capable of keeping a body together. The simplest way to do so is to add in the right-hand side of (4) a second term $-pg_{ik}$, which, for $p > 0$, means (speaking the language of classical mechanics) nothing but an isotropic pressure. Such a pressure, resulting from interior actions, has certainly no Newtonian effect on the motions of the centres of mass of the single bodies. It is not self-evident that the influence of pressure is still zero for the relativistic (first order) correction of these motions.

With the feeling that it must be so, I have omitted from the beginning the term $-pg_{ik}$ in the T'_{ik}, taking only later advantage of the circumstance that the bodies to be considered have, almost sensibly, the characteristics of ordinary solids.

Having now extended the analysis of the question, in order to account for p at every step, I just noted that no modification is needed to the conclusions of the preceding investigation, obtaining thus the expected mathematical support to my early dropping of the term in p, which, however, I willingly recognize, does not appear satisfactory from relativistic point of view.

INSTITUTE FOR ADVANCED STUDY,
PRINCETON UNIVERSITY,
November 2, 1936.

FINITE OVA.[1]

By A. R. Poole.

In the following discussion the term finite ovum or simply ovum is used to denote a system consisting of a finite set of distinct elements and a commutative and associative binary rule of combination. The rule of combination is called multiplication and the customary terminology of products and powers is employed. If the ovum has n elements it is said to be of order n. For each ovum of order n there exists a set of $n(n+1)/2$ relations, the multiplication table of the ovum. Two ova of order n are distinct when they are not simply isomorphic.[2] In ovum theory such terms as zero element, identity element, irreducible element, divisor of an element, proper divisor, associate elements, sub-ovum, have the same significance as they have in ring or field theory.

Powers of elements in an ovum. In the chain of elements obtained by taking successive integral powers

$$u_i, u_i^2, u_i^3 \cdots u_i^q \cdots$$

of an element u_i of an ovum, let the p-th power be the first one which gives an element previously occurring in the chain, and suppose

$$u_i^p = u_i^r$$

Let $p - r = s$, so that

$$u_i^{r+s} = u_i^r.$$

There are four possibilities.

$$\begin{array}{lll} (1) & r = 1, & s = 1 \\ (2) & r > 1, & s = 1 \\ (3) & r = 1, & s > 1 \\ (4) & r > 1, & s > 1. \end{array}$$

We shall refer to r as the index and s as the period of the element u_i. If (1) is true u_i is called an idempotent element and evidently all powers of u_i are equal to u_i. If (2) is true we shall call u_i an element of type A. Then

$$u_i, u_i^2, u_i^3 \cdots u_i^r$$

[1] Received September 1, 1936.

[2] Van der Waerden, *Moderne Algebra*, vol. 1, p. 29.

are distinct and

$$u_i^{r+1} = u_i^r.$$

Hence all powers greater than r of u_i are equal to u_i^r. If $u_i^r = u_k$ we shall call u_k the index element of u_i. The element u_k is an idempotent element and is moreover the only idempotent element occurring in the chain of powers of u_i. The elements

$$u_i^t \qquad (t = 1, \cdots, r-1)$$

are elements of type A with indices less than r. If (3) is true we shall call u_i an element of type B. In this case

$$u_i, u_i^2, \cdots, u_i^s$$

are distinct, and

$$u_i^{s+1} = u_i.$$

Evidently

$$u_i^{s+t} = u_i^t \qquad (t \geqq 1)$$

and in particular

$$u_i^{2s} = u_i^s$$

so we see that u_i^s is an idempotent element and if

$$u_i^s = u_k$$

we call u_k the period element of u_i. It is easily shown that u_k is the only idempotent element occurring in the chain of powers of u_i and moreover that every element

$$u_i^t, \qquad (t < s)$$

is an element of type B and has the same period element as u_i. If (4) is true we shall call u_i an element of type C. In this case

$$u_i, u_i^2, \cdots, u_i^r, \cdots, u_i^{r+s-1}$$

are distinct and

$$u_i^{r+s} = u_i^r$$

whence for an integer $p \geqq r$ and h any integer

$$u_i^{p+hs} = u_i^p.$$

The elements

$$u_i, u_i^2 \cdots u_i^{r-1}$$

form the unrepeated part of the chain of powers of u_i, the elements

$$u_i^r, \cdots, u_i^{r+s-1}$$

form the repeated part of the chain. Let m be the least integer such that $ms \geqq r$. Then u_i^{ms} is an idempotent element and if

$$u_i^{ms} = u_k$$

then u_k is called the period element of u_i. The element u_k is the only idempotent element occurring in the chain of powers of u_i and it can easily be shown that of the elements occurring in the unrepeated part of the chain those being powers of u_i which are multiples of s are type A elements and the others are of type C, while those other than the period element in the repeated part of the chain are type B elements.

We see that each non-idempotent element in an ovum has one and only one idempotent element in its chain of powers. We shall call it the idempotent element of the non-idempotent element.

As an immediate consequence of the preceding we can state

THEOREM 1. *Every ovum contains at least one idempotent element.*

Homomorphisms[3] *in ova containing non-idempotent elements.* In an ovum which contains at least one non-idempotent element, consider the correspondence formed by letting each idempotent element correspond to itself and each non-idempotent element correspond to its idempotent element. Then as such a correspondence is preserved under multiplication we can state

THEOREM 2. *Any ovum which contains at least one non-idempotent element is homomorphic to the sub-ovum formed by its idempotent elements.*

Associate elements in ova. The following theorems on associate elements in ova are easily verified.

THEOREM 3. *In any ovum no two idempotent elements can be associated.*

THEOREM 4. *In any ovum no two non-idempotent elements which have not the same idempotent element can be associated.*

THEOREM 5. *In any ovum a non-idempotent element can not be associated to an idempatent element which is not its idempotent element.*

THEOREM 6. *In any ovum no type A or type C element can be associated*

[3] Van der Waerden, *Moderne Algebra*, p. 32.

to its idempotent element, but every type B element is associated to its idempotent element.

THEOREM 7. *In any ovum no type A or type C element can be associated to an idempotent element or to an element of type B.*

THEOREM 8. *In any ovum two type B elements are associated if and only if they have the same period element.*

THEOREM 9. *In any ovum two type A elements, two type C elements, or an element of type A and an element of type C, can not be associated if they do not have the same index.*

Reduced ova. An ovum in which no pair of elements are associated to one another is called a reduced ovum. From Theorem 6 it follows that a reduced ovum can contain no elements of type B and consequently no elements of type C, and thus contains only elements of type A and idempotent elements We give in detail the proof of the converse of this, namely

THEOREM 10. *Every ovum which contains only idempotent elements and elements of type A is reduced.*

Theorems 3, 4, 7, and 9 show that this result will follow if we prove that in an ovum which contains only idempotent elements and elements of type A, no two type A elements which have the same index element and the same index can be associated. To do this we make use of the following lemmas which are easily verified.

LEMMA 1. *In any ovum if u_i is a type A element with index element u_k, and u_m is another idempotent element such that*

$$u_k u_m = u_m$$

then

$$u_i u_m = u_m.$$

LEMMA 2. *In any ovum, if u_i and u_j are two type A elements with the same index element u_k and the same index, then none of the relations*

$$u_i u_k = u_i$$
$$u_i u_k = u_j$$
$$u_i u_j = u_j$$
$$u_i^2 = u_j$$

are possible.

Returning to the main theorem, let u_i and u_j be two type A elements with index element u_k and index r, in an ovum 0 which contains only idempotent elements and elements of type A. Then

(1) $$u_i u_k = u_k$$
(2) $$u_j u_k = u_k.$$
Assume
(3) $$u_i \sim u_j.$$ [4]

Then there exists in 0 an element whose product with u_i is equal to u_j, and by Lemma 2 this element is neither u_i, u_j, or u_k. There must therefore exist in 0 another element say u_l which is such that

(4) $$u_i u_l = u_j.$$

There must also exist in 0 an element whose product with u_j is equal to u_i. From Lemma 2 this element is neither u_i, u_j, or u_k. We show that it cannot be u_l. Assuming
(5) $$u_j u_l = u_i$$

and combining this with (4), we have

(6) $$u_j u_l^2 = u_j$$

(7) $$u_i u_l^2 = u_i.$$
Now
$$u_l^2 = u_l$$
would imply from (6)
$$u_j u_l = u_j$$

in contradiction to (5), so that under assumption (5) u_l^2 cannot equal u_l. From Lemma 2 and (6) and (7) it follows that u_l^2 cannot equal any of u_i, u_j, or u_k. Hence if (5) holds 0 must contain another element u_m such that

(8) $$u_l^2 = u_m$$
(9) $$u_j u_m = u_j$$
(10) $$u_i u_m = u_i.$$

From (8) u_m is either the index element of u_l or is a type A element having the same index element as u_l. Hence for s sufficiently large (i. e. $s \geqq 1$ if u_m is idempotent, $s \geqq t$ where t is the index of u_m if u_m is a type A element)

---

[4] u_i is associated to u_j.

(11) $$u_l u_m^s = u_m^s.$$
From (9) and (10) we have

(12) $$u_j u_m^s = u_j u_m^{s-1} = u_j u_m = u_j,$$
and
(13) $$u_i u_m^s = u_i u_m^{s-1} = u_i u_m = u_i.$$

Multiplying both sides of (4) by u_m^s gives

$$u_i u_l u_m^s = u_j u_m^s$$

which on employing (11) and (12) gives

$$u_i u_m^s = u_j$$

in contradiction to (13). Thus (5) is impossible; so 0 must contain besides u_i, u_j, u_k, and u_l an element u_n which is such that

(14) $$u_j u_n = u_i.$$
From (4) and (14)

$$u_i u_n u_l = u_i$$
$$u_j u_n u_l = u_j.$$

From Lemma 2, $u_n u_l$ can equal neither u_k, u_i, nor u_j; from (4) it cannot equal u_l and from (14) it cannot equal u_n. Hence 0 contains another element u_p which is such that

(15) $$u_n u_l = u_p$$
(16) $$u_i u_p = u_i$$
(17) $$u_j u_p = u_j.$$

We show now that of the three elements u_n, u_p, and u_l no one can be the index element of any other, no pair of them can have the same index element and no one of them has index element u_k. For if u_p is the index element of u_n or if u_n is the index element of u_p or if u_n and u_p both have the same index element, for t sufficiently large ($t \geqq 1$ if u_p is an idempotent element or t greater than or equal to the index of u_p if u_p is a type A element)

(18) $$u_n u_p^t = u_p^t.$$

Moreover from (16) and (17) we have

(19) $$u_i u_p^t = u_i$$
(20) $$u_j u_p^t = u_j$$
and from (14)

$$u_j u_n u_p{}^t = u_i u_p{}^t$$

which using (18) and (19) gives

$$u_j u_p{}^t = u_i$$

in contradiction to (20). Similarly u_p cannot be the index element of u_l, u_l cannot be the index element of u_p and u_p and u_l cannot have the same index element. Also if u_n and u_l had the same index element, u_p would be that index element or would have the same index element by (15), in contradiction to what we have just shown. If u_n were the index element of u_l or u_l were the index element of u_n we would have

$$u_n u_l = u_n$$

or

$$u_n u_l = u_l$$

in contradiction to (15).

To show that no one of u_l, u_n or u_p has index element u_k, we first assume that u_p has index element u_k, so that for t greater than or equal to the index of u_p

$$u_p{}^t = u_k$$

whence from (1)

$$u_i u_p{}^t = u_k$$

in contradiction to (19); hence u_p cannot have u_k as its index element. Assume that u_l has index element u_k; then by the above u_n does not have index element u_k, and does not have index element u_p. So u_n is either an idempotent element or there exists in 0 another element u_q which is the index element of u_n. If u_n is an idempotent element, from (14) and Theorem 2, follows

(21) $$u_k u_n = u_k$$

otherwise for u_n a type A element, from (14) and Theorem 2

(22) $$u_k u_q = u_k.$$

From (15) for any integer s

$$u_n{}^s u_l{}^s = u_p{}^s$$

and for s sufficiently large

(23) $$u_n u_k = u_p{}^s$$

if u_n is an idempotent element, otherwise

(24) $$u_q u_k = u_p{}^s.$$

For u_n an idempotent element combining (21) and (23) gives

$$(25) \qquad u_p^s = u_k$$

and for u_n a type A element combining (22) and (24) we get (25). But (25) indicates that u_k is the index element of u_p which we have proved impossible. Hence u_l and similarly u_n cannot have u_k as its index element. We have now shown that assumption (3) implies that 0 contains besides the three elements u_i, u_j, and u_k, at least three more elements u_l, u_n, and u_p, and that these six elements satisfy the seven relations

$$(1) \qquad u_i u_k = u_k$$
$$(2) \qquad u_j u_k = u_k$$
$$(4) \qquad u_i u_l = u_j$$
$$(14) \qquad u_j u_n = u_i$$
$$(15) \qquad u_l u_n = u_p$$
$$(16) \qquad u_i u_p = u_i$$
$$(17) \qquad u_j u_p = u_j.$$

Moreover of the three elements u_l, u_n, and u_p, no one can be the index element of any other, no pair of them can have the same index element, and no one of them can have index element u_k. Now let u_L denote the index element of u_l if u_l is a type A element, and let u_L denote u_l itself if u_l is idempotent. Let u_N and u_P have similar significance with respect to u_n and u_p. Then from (15) and Theorem 2 follows

$$(26) \qquad u_L u_N = u_P.$$

Multiplying both sides of (26) by u_L gives

$$(27) \qquad u_P u_L = u_P$$

which by Lemma 1 gives

$$(28) \qquad u_P u_l = u_P.$$

From (16) for any integer t

$$u_p{}^t u_i = u_i$$

which for t sufficiently large gives

$$(29) \qquad u_P u_i = u_i.$$

Similarly from (17)

$$(30) \qquad u_P u_j = u_j.$$

From (4) multiplying both sides by u_P

$$u_i u_l u_P = u_j u_P$$

which on employing (28) and (30) yields

$$u_i u_P = u_j$$

in contradiction to (29). Thus assumption (3) leads to a contradiction so that we conclude that in an ovum which contains only idempotent elements and elements of type A no type A elements having the same index element and the same index can be associated. Thus the proof of Theorem 10 is now complete.

Further results for reduced ova which are easily verified are:

THEOREM 11. *A reduced ovum must contain a zero element.*

THEOREM 12. *If a reduced ovum has an identity element, that element is an irreducible element.*

COROLLARY 1. *A reduced ovum O containing an identity element has a reduced sub ovum of order $n-1$, consisting of all the elements of O except the identity.*

COROLLARY 2. *From a reduced ovum O of order n mark set $(u_1, u_2 \cdots u_n)$, we can form a reduced ovum of order $n+1$ containing an identity element, by adjoining to the mark set of O an element u_{n+1} and to the multiplication table of O the relations*

$$u^2_{n+1} = u_{n+1}$$
$$u_{n+1} u_i = u_i u_{n+1} = u_i. \qquad (i = 1, 2, 3 \cdots n).$$

THEOREM 13. *A reduced ovum has at least one irreducible element which is not an identity element.*

COROLLARY 1. *Every reduced ovum of order n has at least one reduced sub ovum of order $n-1$.*

This corollary shows that from all possible distinct ova of order $n-1$ we can obtain all possible distinct ova of order n by adjoining to the ova of order $n-1$ another idempotent or type A element making multiplication of this element with itself and with the original elements commutative and associative, and examining the ova thus formed to see which are simply isomorphic to one another.

Ova formed from groups. We consider now those ova which contain no

type A or type C elements but at least one element of type B. For convenience we call these ova of type 2.

If G is a finite abelian group of order $n > 1$ and i is its identity element every element a of G has the property

$$a^n = i$$

whence

$$a^{n+1} = a$$

and so G is an ovum of type 2 with only one idempotent element the identity. Conversely an ovum O of type 2 containing only one idempotent element is a group. More generally we can state

THEOREM 14. *Every ovum O of type 2 is either a group, or consists of sub ova which have no element in common and each of which is a group. Each of these groups consists of an idempotent element and of all the type B elements which have this idempotent element for period element.*

In a group every element divides every other element so that every element is associated to every other element. From Theorems 3 and 6 it follows that an ovum in which every element is associated to every other element can have only one idempotent element and type B elements and is therefore a group. Hence we have

THEOREM 15. *A sufficient condition that an ovum be a finite group is that every element be associated to every other element.*

It must be noted that the condition given in this theorem for finite abelian groups is not as strong as that given by Van Der Waerden in his postulates for groups.[5] His postulate 5 not only demands that every element be associated to every other element, but that every element divide itself.

From Theorem 15 we know that all type 2 ova of order n can be obtained by compounding groups of order $\leq n$, only those combinations of groups being taken the sum of whose orders is equal to n. The commutative and associative laws must be satisfied and the ova must be tested for distinctness. Of interest in the question of forming type 2 ova from groups is

THEOREM 16. *From a finite abelian group of order $n - 1$ we can obtain two and only two ova of order n, by the adjunction of an idempotent element.*

MINNEAPOLIS, MINN.

[5] Van der Waerden, *Moderne Algebra*, p. 19, (5).

THE INTEGERS REPRESENTED BY SETS OF POSITIVE TERNARY QUADRATIC NON-CLASSIC FORMS.[1]

By O. K. Sagen.

Introduction. A. A. Albert[2] has shown for the classic case that all integers which are not represented by a *set* of positive ternary quadratic forms lie in certain arithmetic progressions. In this paper analogous results are obtained for the non-classic forms by methods similar to Albert's. The problem is solved by determining under what conditions there exist solutions of a Diophantine equation of certain form.

1. *Proper representation by sets.* Consider the positive ternary quadratic non-classic form

$$(1) \qquad \begin{aligned} &f(x, y, z) = ax^2 + by^2 + cz^2 + ryz + szx + txy \\ &a > 0,\ b > 0,\ c > 0,\ r,\ s,\ t \text{ not all even, } 4bc - r^2 > 0. \end{aligned}$$

The determinant of the form $2f(x, y, z)$ is

$$2d = 2(4abc + rst - ar^2 - bs^2 - ct^2).$$

Since the determinant $2d$ is an invariant of the form under unitary transformation we shall define the *set* $\Sigma(d)$ as the set of all forms (1) having the same invariant d. The set $\Sigma(d)$ is said to *represent* an integer m, if there exist integers x, y, z and a form f of invariant d such that $m = f(x, y, z)$. If the g. c. d. of x, y, z is unity, then the representation is called *proper*.

2. *Necessary and sufficient conditions for proper representation.* Suppose that an integer a is properly represented by a set $\Sigma(d)$. It is well known[3] that if an integer a is properly represented by some form, there exists an equivalent form f having a as the coefficient of x^2. Since classic forms are non-equivalent to non-classic forms and since equivalent forms represent the same integers, there exists a positive non-classic form with coefficients a, b, c, r, s, t such that

$$(3) \quad d + ar^2 + bs^2 + ct^2 - rst = 4abc,\ 4bc - r^2 > 0 \ (r, s, t \text{ not all even}).$$

Conversely, if (3) holds then the integer a is properly represented by $\Sigma(d)$.

[1] Received January 22, 1934; revised October 15, 1936.

[2] A. A. Albert, "The integers represented by sets of ternary quadratic forms," *American Journal of Mathematics*, vol. 55 (1933), pp. 274-292.

[3] L. E. Dickson, *Studies in the theory of numbers*, 1930. University of Chicago Press, p. 12.

LEMMA. *An integer a is properly represented by a set $\Sigma(d)$, if and only if there exist integers b, c, r, s, t such that (3) holds.*

Suppose that (3) is true for given integers a and d. The binary quadratic form $\phi(y, z) = by^2 + ryz + cz^2$ has discriminant $-\Delta = r^2 - 4bc < 0$ and is a positive form, since $b > 0$. By (3), evidently

$$\phi(s, -t) = bs^2 + ct^2 - rst = 4abc - ar^2 - d = a\Delta - d. \tag{4}$$

Let $s_1 =$ g. c. d. (s, t). Then, on writing $s = s_1 s_0$, $t = s_1 t_0$, $b_1 = \phi(s_0, -t_0)$, we have

$$\phi(s, -t) = s_1^2 b_1. \tag{5}$$

Since ϕ properly represents the positive integer b_1, there exists an equivalent binary ϕ_1 of discriminant $-\Delta$ such that

$$\phi_1(y_1, z_1) = b_1 y_1^2 + r_1 y_1 z_1 + c_1 z_1^2.$$

Therefore there exist integers which satisfy

$$\Delta = 4b_1 c_1 - r_1^2. \tag{6}$$

But, by (4), $b_1 s_1^2 = a\Delta - d$, and consequently (6) implies the existence of integers b_1, c_1, r_1, s_1 such that

$$d + ar_1^2 = b_1(4ac_1 - s_1^2) > 0. \tag{7}$$

By (6) and (4), $r^2 \equiv -\Delta \equiv r_1^2 \pmod 4$; so that, since $s_1 =$ g. c. d. (s, t), not both r_1 and s_1 are even. Evidently (7) implies (3); so by the lemma we obtain the following theorem.[4]

THEOREM 1. *A positive integer a is represented properly by a set $\Sigma(d)$, if and only if there exist positive integers b, c, r, s, with r, s not both even, such that*

$$d + ar^2 + bs^2 = 4abc. \tag{8}$$

COROLLARY 1. *If equation (8) holds with s odd, then r is of arbitrary parity.*

COROLLARY 2. *In equation (8) r, s, c may be replaced respectively by $2bv + r$, $2au + s$, $au^2 + bv^2 + c + rv + su$ where u, v are arbitrary integers.*

When (8) is true, evidently

$$d + a(r + s)^2 + (b + a)s^2 + c(2a)^2 - (r + s)s(2a) = 4a(b + a)c. \tag{9}$$

---

[4] This theorem and its proof are valid for sets of classic forms and from them essential simplifications of Albert's proofs are obtainable.

Since (9) is an equation of the same form as (3), the proof of Theorem 1 shows that integers b_1, c_1, r_1, s_1 exist, such that (7) is true and $r_1^2 \equiv (r+s)^2 \pmod 4$. But since s is odd, we have $r_1 \equiv r+1 \pmod 2$, whence Corollary 1 follows.

Writing $r_1 = 2bv + r$, $s_1 = 2au + s$, $c_1 = au^2 + bv^2 + c + rv + su$, we note that

$$d + ar_1^2 + bs_1^2 = 4abc_1$$

reduces to (8); which verifies Corollary 2.

3. *Necessary and sufficient condition for representation.* Suppose that the positive integers a and d are expressed in their unique forms $a = 4^l\lambda^2 M$, $d = 4^g\Delta$, where λ is odd, M is without square factor, and Δ is not divisible by 4.

THEOREM 2. *A positive integer a is represented by a set $\Sigma(d)$ of non-classic forms, if, and only if, there exist positive integers b, c, r, s such that*

(10) $\Delta + Mr^2 + bs^2 = 4Mbc$, *where s is odd whenever $l < g$, while, if $l \geqq g$, at least one of r and s is odd.*

If a is represented by $\Sigma(d)$, there exists a form f and integers ξ, η, ζ such that $f(\xi, \eta, \zeta) = L^2M$. Let $\theta = \text{g. c. d.}\ (\xi, \eta, \zeta)$, $\xi = \theta\xi_1$, $\eta = \theta\eta_1$, $\zeta = \theta\zeta_1$. Then $4^l\lambda^2M = \theta^2 f(\xi_1, \eta_1, \zeta_1)$. Evidently θ is a divisor of $2^l\lambda$, and on writing $2^l\lambda = 2^h\lambda_0\theta$, we note that $4^h\lambda_0^2M$ is properly represented by $\Sigma(d)$. By Theorem 1,

$$4^g\Delta + 4^h\lambda_0^2Mr^2 + bs^2 = 4\cdot 4^h\lambda_0^2Mbc. \tag{11}$$

If s is odd, on setting $r_1 = 2^h\lambda_0 r$ we have

$$4^g\Delta + Mr_1^2 + bs^2 \equiv 0 \pmod{4Mb}. \tag{12}$$

Otherwise, let $s = 2^\sigma s_1$, where s_1 is odd and $\sigma > 0$. We then obtain

$$4^g\Delta + 4^h\lambda_0^2Mr^2 + 4^\sigma bs_1^2 \equiv 0 \bmod (4\cdot 4^hMb).$$

If $\sigma > h$ then, either $g = h$, whence

$$\Delta + M(\lambda_0 r)^2 + b(2^{\sigma-h}s_1)^2 \equiv 0 \pmod{4Mb}, \tag{13}$$

or $g > h$, so that $\sigma = h$, and

$$4^{g-h}\Delta + M(\lambda_0 r)^2 + bs_1^2 \equiv 0 \pmod{4Mb}. \tag{14}$$

If $\sigma \leqq h$, then $g \geqq \sigma$ and, writing $r_2 = 2^{h-\sigma}\lambda_0 r$, we have

$$4^{g-\sigma}\Delta + Mr_2^2 + bs_1^2 \equiv 0 \pmod{4Mb}. \tag{15}$$

But, if $4^k\Delta + Mr_0^2 + bs_0^2 \equiv 0 \pmod{4Mb}$, we may take $r_0 = 2\rho$ in view of Corollary 1, Theorem 1. Hence $b = 4\beta$ and

$$4^{k-1}\Delta + M\rho^2 + \beta s_0^2 \equiv 0 \pmod{4M\beta}.$$

Therefore, (12), (13), (14), (15) imply the existence of integers b^1, r^1, s^1, c^1, such that

(16) $\Delta + Mr^{12} + b^1 s^{12} = 4Mb^1c^1$, where s^1 is even in case $g = h$, while otherwise s^1 is odd. Hence, if $g > l$, (16) holds with s^1 odd; while for $g \leqq l$, (16) holds with at least one of r^1 and s^1 odd.

Conversely, if the conditions (16) hold for s^1 odd, we may write $r = 2^g r^1$, $b = 4^g b^1$; therefore

$$4^g \Delta + Mr^2 + bs^{12} = 4Mbc^1.$$

Thus by Theorem 1, M is properly represented by $\Sigma(d)$. But when s^1 is even, $g = h \leqq l$, and r^1 is odd. Setting $s = 2^g s^1$, we have

$$4^g \Delta + 4^h M r^{12} + b^1 s^2 = 4^{h+1} M b^1 c^1.$$

Therefore $4^h M$ is properly represented by $\Sigma(d)$. By definition then, there exists a form f_1 for s^1 odd such that $f_1(1, 0, 0) = M$, while for s^1 even, a form f_2 exists such that $f_2(1, 0, 0) = 4^h M$. But $f_1(2^l \lambda, 0, 0) = 4^l \lambda^2 M$ and $f_2(2^{l-h}\lambda, 0, 0) = 4^l \lambda^2 M$, so a is therefore represented by $\Sigma(d)$.

4. *Reduction of the general case to a problem for integers without square factors.*

LEMMA 1. *Let Ka be any integer without square factor with K odd, and let d be any integer. There exist solutions b, c, r, s of*

$$K^2 d + Kar^2 + bs^2 = 4Kabc$$

if and only if there exist solutions b^1, c^1, r^1, s^1 of

$$Kd + ar^{12} + b^1 s^{12} = 4ab^1c^1.$$

Furthermore, s and s^1 are of the same parity.

LEMMA 2. *Let p be an odd prime not dividing a such that either the Legendre symbol $(-ad/p) = +1$, or d is divisible by p. There exist solutions of*

$$p^2 d + ar^2 + bs^2 = 4abc$$

if and only if s, s^1 are of the same parity and

$$d + ar^{12} + b^1 s^{12} = 4ab^1c^1$$

is solvable.

LEMMA 3. *Let p be an odd prime such $(ad/p) = -1$. Then*

$$p^2 d + ar^2 + bs^2 = 4abc$$

is satisfied if and only if either

$$pd + par^{1^2} + b^1 s^{1^2} = 4pab^1c^1$$

or

$$d + ar^{1^2} + b^1 s^{1^2} = 4ab^1c^1$$

hold with s^1 of the same parity as s.

For the proof of Lemma 1, let the g. c. d. $(b, K) = \beta$, then $K = \beta K_0$, $b = \beta b_1$, $s = K_0 s^1$. By Corollary 2 of Theorem 1, r may be taken divisible by K_0. Hence

$$K^2 d + Kar^2 + bs^2 \equiv 0 \pmod{4KK_0ab},$$

and, on writing $K_0 b_1 = b^1$, we have

(17) $$Kd + ar^2 + b^1 s^{1^2} \equiv 0 \pmod{4ab^1}.$$

Conversely, if (17) holds, let the g. c. d. $(b^1, K) = K_0$ and $K = K_0 K_1$, $b^1 = K_0 b_1$. By Corollary 2, Theorem 1, r may be chosen divisible by K_1 so that (17) holds modulo $4ab^1K$. Thus, setting $s = K_0 s^1$, $b = b^1 K$, and multiplying (17) by K, we obtain

$$K^2 d + Kar^2 + bs^2 \equiv 0 \pmod{4Kab}.$$

Evidently $s \equiv s^1 \pmod 2$, and, hence, Lemma 1 follows.

We prove Lemmas 2 and 3 simultaneously. In case b is not divisible by p, we may, by Corollary 2, Theorem 1, take $r = pr_1$ and $s = ps_1$ so that

(18) $$d + ar_1^2 + bs_1^2 \equiv 0 \pmod{4ab}.$$

But if $b = pb_1$, then $r = pr_1$, and

(19) $$pd + par_1^2 + b_1 s^2 \equiv 0 \pmod{4pab_1}.$$

If $b_1 = pb_2$, then

(20) $$d + ar_1^2 + b_2 s^2 \equiv 0 \pmod{4ab_2}.$$

If b_1 is not divisible by p, then in view of Corollary 2, Theorem 1, r_1 may be chosen so that $d + ar_1^2 \equiv 0 \pmod p$, whenever either p divides d or $(-ad/p) = +1$. Then (19) holds modulo $4p^2ab_1$; therefore,

(21) $$d + ar_1^2 + bs_1^2 \equiv 0 \pmod{4ab}.$$

Conversely, if any of (18), (20), (21) hold, then on multiplying by p^2 we obtain

$$p^2 d + ar^2 + b^1 s^2 \equiv 0 \pmod{4ab^1}.$$

If (19) holds, then, on multiplying by p, we have

$$p^2 d + ar^2 + bs^2 \equiv 0 \pmod{4ab}.$$

Evidently $s \equiv s_1 \pmod 2$; hence the lemmas are verified.

Now express Δ of Theorem 2 in its unique form $\Delta = \gamma^2 D$ where γ is odd and D is without square factor. Let $\mu =$ g. c. d. (M, γ), $M = \mu M_1$, $\gamma = \mu\gamma_1$. By Lemma 1, condition (10) holds if and only if it also holds for Δ and M replaced by $\gamma_1{}^2\mu D$ and M_1 respectively. Now let the g. c. d. $(\mu, D) = \mu_0$, $\mu = \mu_0\mu_1$, $D = \mu_0 D_1$, and let P be the product of all primes p dividing $\gamma_1\mu_0$ such that $(-M_1\mu_1 D_1/p) = -1$. Then, by Lemma 2, condition (10) holds if and only if it also holds with Δ, M replaced by $P^2\mu_1 D_1$, M_1 respectively. Finally, by Lemma 3, (10) holds if and only if it holds either for the pair $\mu_1 D_1$, M_1 or for $P\mu_1 D_1$, PM_1.

THEOREM 3. *Let* $a = 4^l\lambda^2 M$, $d = 4^g\gamma^2 D$, *where* λ, γ *are odd,* M *and* D *each without square factor. Further suppose* $\mu_0 =$ g. c. d. (M, D, γ), $\mu_0\mu_1 =$ g. c. d. (M, γ), $M = \mu_0\mu_1 M_1$, $D = \mu_0 D_1$, $\gamma = \mu_0\mu_1\gamma_1$, *and let* P *be the product of all primes* p *dividing* $\mu_0\gamma_1$ *for which* $(-M_1\mu_1 D_1/p) = -1$. *Then* a *is represented by* $\Sigma(d)$ *if and only if there exist integers* b, c, r, s *such that either*

$$\mu_1 D_1 + M_1 r^2 + bs^2 = 4M_1 bc \quad \textit{or} \quad P\mu_1 D_1 + PM_1 r^2 + bs^2 = 4PM_1 bc$$

with s *odd, if* $l < g$, *and with* r, s *not both even, if* $l \geqq g$.

5. *A sufficient condition.*

THEOREM 4. *If* $\mu_1 > 1$ *or if* D_1 *has a factor prime to* M_1, *then* a *is represented by* $\Sigma(d)$ *where* M_1 *and* D_1 *are defined as in Theorem* 3.

For the proof, let $\Delta = 2^i q\theta$, $N = 2^j\mu$ each be an integer without square factor, such that q is prime to N, and $\mu\theta$ is odd.

First, suppose q is odd and define

$$p_n = [Bq]n + [B^{q-1}(w+1) - 1], \tag{23}$$

such that $B = 8\mu\theta$ and $(w/q) = -(-1/q)$. Since B and q are relatively prime, $B^{q-1} \equiv 1 \pmod q$, by Fermat's theorem. Consequently $p_n \equiv w \not\equiv 0 \pmod q$, $p_n \equiv -1 \pmod B$, and the bracketed expressions in (23) are relatively prime. By Dirichlet's theorem [5] on primes in arithmetic progression p_n represents

[5] P. G. L. Dirichlet, *Abhandlungen der Königlichen Akademie der Wissenschaften*, Berlin, 1837, pp. 108-110.

an infinitude of primes. If, for $n = m$, p_m is an odd prime, write $p_m = p$. We can now compute the Legendre symbol $(-N\Delta/p) = (-2^{i+j}q\theta\mu/p)$. First, $(2/p) = +1$ and $(-1/p) = -1$, since $p \equiv -1 \pmod 8$. Also $(\mu/p) = (\theta/p) = +1$, since $p \equiv -1 \pmod{\theta\mu}$. But $p \equiv w \pmod q$ and hence

$$(q/p) = (-1)^{(p-1)/2 \, (q-1)/2}(w/q) = -1.$$

Therefore, $(-N\Delta/p) = +1$, and $(-Q\Delta/p) = +1$, where $NQ \equiv 1 \pmod p$. Consequently, there exists an odd integer r such that $r^2 \equiv -Q\Delta \pmod p$. As a result

(24) $$\Delta + Nr^2 \equiv 0 \pmod p.$$

Writing $s = 1$, $c = 2\theta[qm + B^{q-2}(w+1)]$, we have $p = 4Mc - s^2$.

Now, suppose $q = 2$. Then $\Delta = 2\theta$ and $N = \mu$; hence we write

$$p_n = 8\mu\theta n + 4\mu\theta - 1.$$

By Dirichlet's theorem p_n is an odd prime p, by proper choice of n. Evidently $p \equiv 3 \pmod 8$ and $p \equiv -1 \pmod{\mu\theta}$, and, therefore, $(-1/p) = (2/p) = -1$, $(\mu\theta/p) = +1$. Thus the Legendre symbol $(-N\Delta/p) = +1$, so there exists an odd integer r such that

(25) $$\Delta + Nr^2 \equiv 0 \pmod p.$$

But $p = 4Mc - s^2$, where $c = \theta(2n+1)$ and $s = 1$.

Now if $\mu_1 > 1$ in Theorem 3, or if there exists some factor of D_1 which is prime to M_1, then (24) and (25) imply (22) is satisfied with s odd. Therefore a is represented by $\Sigma(d)$.

6. *Necessary and sufficient conditions for proper representation.*

THEOREM 5. *Suppose AD is without square factor. There exist integers b, c, r, s, with s odd, such that*

(26) $$D + ADr^2 + bs^2 = 4ADbc$$

if and only if there exists an odd integer $q > 1$ which divides D and is such that $(-A/q) = -1$.

Suppose (26) holds and let $q = \text{g.c.d.}(s, D)$, $s = qs_0$, $D = qD_0$, $m = 4AD_0c - qs_0^2$. Evidently m is prime to D_0 and hence, by (26),

(27) $$1 + Ar^2 \equiv 0 \pmod m.$$

Therefore, if $m > 1$, the Jacobi symbol $(-A/m) = +1$. Writing $A = 2^i\alpha$,

where α is odd and i is either 0 or 1, we observe that $(2/m)=(2/q)$ when $i=1$, since $m\equiv -q \pmod 8$. Also $(-1/m)=-(-1/q)$ and

$$\begin{aligned}(\alpha/m) &= (-1)^{(\alpha-1)/2\,(m-1)/2}\,(m/\alpha)\\ &= (-1)^{(\alpha-1)/2\,(q+1)/2+(\alpha-1)/2+(\alpha-1)/2\,(q-1)/2}\,(\alpha/q)=+(\alpha/q),\end{aligned}$$

since $m\equiv -qs_0^2 \pmod{4\alpha}$. Consequently,

$$(-A/q)=-(-A/m)=-1.$$

If $m=1$, we have $4AD_0c=qs_0^2+1$ and, therefore, $(-q/\alpha)=+1$, since $(qs_0)^2\equiv -q \pmod\alpha$. But $(-1/q)=-1$, since $q\equiv -1 \pmod 4$, and hence,

$$(-\alpha/q)=(-1)^{1+(\alpha-1)/2\,(q-1)/2+(\alpha-1)/2}(-q/\alpha)=-1.$$

If $i=1$, evidently $q\equiv -1 \pmod 8$ and $(2/q)=1$. Therefore, again $(-A/q)=-1$.

Conversely, if $D=qD_0$, q odd, $(-A/q)=-1$, we form the expression

$$p_n=8AD_0n-q. \tag{28}$$

By Dirichlet's theorem, for some value of n, p_n is a prime p. But, since $p\equiv -q \pmod{8A}$, we have $(-1/p)=-(-1/q)$, $(2/p)=(2/q)$ and $(\alpha/p)=(\alpha/q)$. Therefore, the Legendre symbol

$$(-AD_0^2/p)=-(-A/q)=+1,$$

and so there exists an integer r, such that

$$D_0+AD_0r^2\equiv 0 \pmod p, \qquad D+ADr^2\equiv 0 \pmod{qp}. \tag{29}$$

Setting $c=2n$, $s=q$, evidently $qp=4ADc-s^2$, and hence (29) implies (26).

THEOREM 6. *If AD is without square factor, $D=2\delta$, δ odd and $A\equiv 1 \pmod 4$ or $A\equiv 3 \pmod 8$, then there exist integers b, c, r, s such that r is odd, s even, and*

$$D+ADr^2+bs^2=4ADbc. \tag{30}$$

For the proof when $A\equiv 3 \pmod 8$, let

$$p_n=ADn+A\delta-2.$$

Evidently AD is relatively prime to $A\delta-2$; hence, by Dirichlet's theorem, there is some value of n for which p_n is a prime p. Since $p\equiv -2 \pmod A$, we find

$$(-A/p)=(-1)^{(p-1)/2+(p-1)/2\,(A-1)/2}(-2/A)=+1.$$

Therefore, there exists an odd integer r, such that

$$1 + Ar^2 \equiv 0 \pmod{p}.$$

But, since $A \equiv 3 \pmod{8}$,

$$(31) \qquad D + ADr^2 \equiv 0 \pmod{8p}.$$

Now, with $c = 2n + 1$, $s = 4$, evidently $p = 4ADc - s^2$, and therefore (31) implies (30).

When $A \equiv 1 \pmod{4}$, let $p_n = 2ADn + AD - 1$. For some value of n, by Dirichlet's theorem, p_n is a prime p. Since $p \equiv 1 \pmod{4}$ and $p \equiv -1 \pmod{A}$, we have $(-A/p) = +1$. Therefore, there exists an odd integer r, such that

$$(32) \qquad 1 + Ar^2 \equiv 0 \pmod{p}, \qquad D + ADr^2 \equiv 0 \pmod{4p}.$$

On setting $c = 2n + 1$, $s = 2$, we note that $4p = 4Mc - s^2$, and therefore (32) implies (30).

THEOREM 7. *Let AD be without square factor and let there exist integers b, c, r, s with r odd and s even, such that*

$$(33) \qquad D + ADr^2 + bs^2 = 4ADbc.$$

Then A is odd, and if D is odd, $A \equiv 3 \pmod{4}$ and there exists a factor $q > 1$ dividing D, such that $(-A/q) = -1$. If D is even and $A \equiv 7 \pmod{8}$, there exists a factor $q > 1$ dividing D, such that $(-A/q) = -1$.

Let $D = 2^i\delta$, δ odd, $i = 0$ or 1, $s = 2s_1$, $m_0 = ADc - s_1^2$, and suppose $q = \text{g. c. d.}(\delta, m_0)$, $\delta = q\delta_0$, $m_0 = qm$, $s_1 = qs_0$. Then $4ADc - s^2 = 4qm$, where m is relatively prime to δ and by (33),

$$(34) \qquad 2^i\delta_0(1 + Ar^2) \equiv 0 \pmod{4m}.$$

Write $m = 2^j\mu$, where μ is odd. Since m is prime to δ_0, evidently, for $\mu > 1$, $1 + Ar^2 \equiv 0 \pmod{\mu}$ and $(-A/\mu) = +1$. Since $A \equiv 3 \pmod{4}$, $(A/\mu) = (-1)^{(\mu-1)/2} \cdot (\mu/A)$, and $(-1/A) = -1$. Since $m = 2^j\mu \equiv -q \pmod{A}$, we have

$$(\mu/A) = (2^j/A)(m/A)$$
$$= (-1)^{(q+1)/2}(-1/q)(2^j/A)(-A/q) = -(2^j/A)(-A/q).$$

Hence,

$$(-A/q) = -(2^j/A)(-1)^{(\mu-1)/2}(-1/\mu)(-A/\mu) = -(2^j/A).$$

But, if $A \equiv 7 \pmod 8$, then $(2/A) = +1$; and if $i = 0$, $A \equiv 3 \pmod 8$, then, by (34), m is odd, since $\delta_0(1 + Ar^2) \equiv 4 \pmod{\mathrm{R}}$. Therefore, $(2^j/A) = +1$ and $(-A/q) = -1$.

Lastly, if $\mu = 1$, then $2^j = 2^i A \delta_0 c - q s_0^2$, so that $-2^j q \equiv (q s_0)^2 \pmod A$; and, therefore, $(-2^j q/A) = +1$. But, since $A \equiv 3 \pmod 4$,

$$(-A/q) = (-1/q)(-1)^{(q-1)/2}(q/A) = -(-q/A) = -(2^j/A) = -1.$$

7. *Summary.* By Theorem 4, equations (22) are satisfied, except, possibly, when $\mu_1 = 1$ and $M_1 = AD_1$. But, by Theorem 5, the second of equations (22), with s odd, holds for every $M_1 = AD_1$ when $P > 1$. Therefore, equations (22) can be satisfied with s odd in all cases except when $M = AD$ and $(-A/p) = +1$ for every odd prime p dividing d. By Theorem 6, equations (22) can be satisfied with s even for $M = AD$ whenever A is odd and incongruent to 7 modulo 8. Theorem 7 implies that, when $A \equiv 7 \pmod 8$, if equations (22) are not satisfied with s odd, then they also fail to hold for s even. Hence, by Theorem 3, we obtain the final theorem.

THEOREM 8. *Express any positive integers a and d in their unique forms*

$$a = 4^l \lambda^2 M, \qquad d = 4^g \gamma^2 D$$

where $\lambda\gamma$ is odd, M and D are without square factor. If $l \geqq g$, the set $\Sigma(d)$ of positive non-classic ternaries represents all integers a except those for which $M = AD$, $A \equiv 7 \pmod 8$ and $(-A/p) = +1$ for every odd prime p dividing d. If $l < g$, the set $\Sigma(d)$ represents all integers a except those for which $M = AD$ and $(-A/p) = +1$ for every odd prime p dividing d.

COROLLARY. *Every set $\Sigma(d)$ represents no integer of the form $4^g D(8dm - 1)$. No set of forms $\Sigma(d)$ represents all positive integers.*

This last result offers new proof of the well known fact that no positive ternary represents all positive integers.

IOWA STATE COLLEGE,
AMES, IOWA.

IDEAL WARING THEOREM FOR THE POLYNOMIAL $m(x^3-x)/6-m(x^2-x)/2+x$.[1]

By Alvin Sugar.

This paper is numbered in sequel to a previous paper by the author[2] in which there was obtained the second known universal Waring theorem for a polynomial with a parameter.[3]

Theorem. *Every positive integer is a sum of* $m+3$ *values of* $P(x)=m(x^3-x)/6+x$ *for non-negative integers* x, *where* $m\geqq 16$.

Since the completion of that paper, Dickson[4] has proved the ideal universal Waring theorem for n-th powers, for $n>6$. This is, then, the third universal Waring theorem for a polynomial with a parameter. In this paper we shall add a fourth theorem to this list. And, furthermore, this result, as were the three preceding, is also an ideal result.

5. *A universal Waring theorem obtained by a transformation.* If we subject

$$(13)\quad f(x)=m(x^3-x)/6-m(x^2-x)/2+x,\ x \text{ integral and } \geqq 0,$$

to the transformation $x=y+1$, we get

$$(14)\qquad f(x)=P(y)+1.$$

From (14) and the above theorem, we can immediately conclude that every positive integer $N>m+3$ is a sum of $m+3$ values of $f(x)$, when $m\geqq 16$. Since every integer $\leqq m+3$ is surely a sum of $m+3$ values (0 and 1 being values of $f(x)$), we have proved the first part of the following theorem.

Theorem 8. *Every positive integer is a sum of* $m+3$ *values of* (13) *for* $m\geqq 16$ *and is a sum of nine values for* $1\leqq m\leqq 6$.[5]

[1] Presented to the Society, August 31, 1936. Received by the Editors, October 19, 1936.

[2] *American Journal of Mathematics*, vol. 58 (1936), pp. 783-790.

[3] The first theorem of this kind was proved by Cauchy, *Oeuvres* (2), vol. 6, pp. 320-353, and had been stated earlier by Fermat. It was only recently that James conjectured the existence of similar theorems for cubic polynomials, *American Journal of Mathematics*, vol. 56 (1934), p. 305.

[4] *American Journal of Mathematics*, vol. 58 (1936), pp. 530-535.

[5] It is known that nine values of $P(x)$ suffice for $1\leqq m\leqq 6$. See Dickson, *Transactions of the American Mathematical Society*, vol. 36 (1934), p. 739, Theorem 12.

This theorem, however, is not an ideal Waring theorem. But, fortunately, we can obtain the ideal Waring theorem for this polynomial.

Since the constant C_1 of Theorem 1 was taken more than 10 greater than the Dickson-Baker-Webber constant, we have

THEOREM 9. *For $m \geqq 7$, every integer $\geqq 10^{12}m^{10}$ is a sum of nine or ten values of* (13) *according as the congruence $m \equiv 6 \pmod 9$ does not or does hold.*

6. *The ideal for $f(x)$.* We list a set of the first twelve values of $f(x)$:

$0, 1, 2, a = m + 3, b = 4m + 4, c = 10m + 5, d = 20m + 6, e = 35m + 7,$ $f = 56m + 8,\ \ g = 84m + 9,\ \ h = 120m + 10,\ \ i = 165m + 11.$

The ideal $g(f)$ is the smallest value of s for which it is true that every positive integer is a sum of s values of f. We can prove

$$B(f) = [(m + 1)/2] + 3 \leqq g(f).$$

For, we see that the integer

$$\left[\frac{4m + 4}{m + 3}\right] (m + 3) - 1$$

requires at least $[(m + 1)/2] + 3$ values of $f(x)$ when $m \geqq 5$. Our next step will be to prove $g(f) \leqq B(f)$ (whence $g(f) = B(f)$) by proving that every positive integer is a sum of $B(f)$ values of $f(x)$.

7. *Ascension theory.* If we replace x by $y + 1$ in (14) we get the following identity in y:

$$f(y + 1) \equiv P(y) + 1. \tag{15}$$

If we write

$$d(a) = f(a + 1) - f(a),$$

we see by (15) that the equality

$$d(a + 1) \equiv F(a) \tag{16}$$

holds identically in a, where $F(a)$, as we recall, was defined to be

$$P(a + 1) - P(a).$$

From (15) it also follows that

$$f(a + 2) = P(a + 1) + 1 > P(a + 1). \tag{17}$$

Now from the analogue for $f(x)$ of Theorem 2, and these two relations, (16) and (17), we see that we can use the following theorems in making ascensions with the polynomial $f(x)$.

THEOREM 10. *Let every integer n, $c < n \leqq g$, be a sum of $k-1$ values of $f(x)$, and let a be an integer $\geqq 0$ for which $F(a) < g - c$. Then every integer N, $c < N \leqq g + P(a+1)$, is a sum of k values of $f(x)$.*

THEOREM 11. *Let every integer n, $c < n \leqq c + pm + q$, be a sum of k values of $f(x)$, and let t be a real number $\geqq 1$ which satisfies the inequality $F(3t+1) < pm + q$. Then every integer N,*

$$c < N \leqq (9/2)t^{2(3/2)^s} m$$

is a sum of $k + s$ values of $f(x)$.

8. *Two Lemmas.*

LEMMA 5. *Every positive integer $\leqq 20m + 6$ is a sum of $B = B(f)$ values of f for $m \geqq 17$.*

LEMMA 6. *Every integer between $20m + 6$ and $217m + 32$ is a sum of B values of f for $m \geqq 36$.*

In proving Lemma 5 we proceed by a method analogous to that employed in obtaining Lemma 2 and get the following set of numbers, so constructed that between each consecutive pair, B values suffice.

$3a = 3m + 9$, $b = 4m + 4$, $a + b = 5m + 7$, $2a + b = 6m + 10$, $6a + m - 7 = 7m + 11$, $6a + m - 6 = 7m + 12$, $3a + b = 7m + 13$, $2b = 8m + 8$, $a + 2b = 9m + 11$, $c = 10m + 5$, $a + c = 11m + 8$, $2a + c = 12m + 11$, $4a + 2b + m - 8 = 13m + 12$, $3a + c = 13m + 14$, $b + c = 14m + 9$, $a + b + c = 15m + 12$, $3a + 3b + m - 8 = 16m + 13$, $4b = 16m + 16$, $8a + 2b + m - 17 = 17m + 15$, $3a + b + c = 17m + 18$, $2b + c = 18m + 13$, $8a + c + m - 15 = 19m + 14$, $a + 2b + c = 19m + 16$, $d = 20m + 6$.

In order to prove Lemma 6 we construct the following set of intervals, over each of which $[(m-1)/2]$ values suffice. They will overlap for $m \geqq 36$. Hence for $m \geqq 36$ we can conclude that every integer n,

$$20m + 5 < n \leqq 23m + 12, \tag{18}$$

is a sum of $[(m-1)/2]$ values of f.

$(d = 20m + 6,\ 21m + 3)$, $(5b = 20m + 20,\ 21m + 9)$, $(a + d = 21m + 9,\ 22m + 4)$, $(5a + 4b = 21m + 31,\ 22m + 12)$, $(2a + d = 22m + 12, 23m + 5)$, $(6a + 4b = 22m + 34,\ 23m + 13)$.

We note that the interval (18) has been so selected that an ascension is unnecessary; for, in virtue of Theorems 10 and 11 (since Theorems 10 and 11 for $f(x)$ are identical with Theorems 2 and 4 for $P(x)$), we see that the ascension made in establishing Lemma 3 from (6) is valid here. And in the future, whenever an ascension is necessary we shall be careful to select our interval and the number of values which suffice over this interval in such a way that a new ascension will not be necessary i. e. in such a way that a corresponding ascension of the previous paper will suffice.

9. *The major ascensions.* The following intervals have been so constructed that every integer lying in anyone of them is a sum of $B - 11$ values of f. These intervals will overlap for $m \geqq 36$. Hence $B - 11$ values will suffice from $165m + 11$ to $171m + 14$.

$(i = 165m + 11,\ 166m - 6)$, $(c + e + h = 165m + 22,\ 166m + 1)$, $(b + 3e + f = 165m + 33,\ 166m + 8)$, $(a + 4d + g = 165m + 36,\ 166m + 9)$, $(a + 4c + 2d + g = 165m + 44, 166m + 13)$, $(a + b + 2c + 4e = 165m + 45,\ 166m + 14)$, $(a + i = 166m + 14,\ 167m - 5)$, $(a + c + e + h = 166m + 25,\ 167m + 2)$, $(a + b + 3e + f = 166m + 36,\ 167m + 9)$, $(2a + 4d + g = 166m + 39,\ 167m + 10)$, $(2a + b + d + 4e = 166m + 44,\ 167m + 13)$, $(a + b + 3c + 2d + e + f = 166m + 49,\ 167m + 16)$, $(a + 4c + d + 3e = 166m + 50,\ 167m + 17)$, $(2a + i = 167m + 17,\ 168m - 4)$, $(2a + c + e + h = 167m + 28,\ 168m + 3)$, $(2a + b + 3e + f = 167m + 39, 168m + 10)$, $(3a + 4d + g = 167m + 42,\ 168m + 11)$, $(3a + b + d + 4e = 167m + 47,\ 168m + 14)$, $(3a + 4c + 2d + g = 167m + 50,\ 168m + 15)$, $(3a + b + 2c + 4e = 167m + 51,\ 168m + 16)$, $(2a + b + 3c + 2d + e + f = 167m + 52,\ 168m + 17)$, $(2a + 4c + d + 3e = 167m + 53,\ 168m + 18)$, $(2g = 168m + 18,\ 169m - 1)$, $(3a + c + e + h = 168m + 31,\ 169m + 4)$, $(2b + 2c + f + g = 168m + 35,\ 169m + 8)$, $(a + 2b + 2c + d + e + g = 168m + 43,\ 169m + 12)$, $(a + 2b + 4c + e + g = 168m + 47,\ 169m + 14)$, $(4a + b + d + 4e = 168m + 50,\ 169m + 15)$, $(b + i = 169m + 15,\ 170m - 4)$, $(b + c + e + h = 169m + 26,\ 170m + 3)$, $(2b + 3e + f = 169m + 37, 170m + 10)$, $(a + 2b + d + 4e = 169m + 45,\ 170m + 14)$, $(2b + 3c + 2d + e + f = 169m + 50,\ 170m + 17)$, $(b + c + 6d + e = 169m + 52,\ 170m + 18)$, $(a + b + i$

$= 170m + 18,\ 171m - 3)$, $(a + b + c + e + h = 170m + 29,\ 171m + 4)$, $(a + 2b + 3e + f = 170m + 40,\ 171m + 11)$, $(2a + b + 2c + 3d + g = 170m + 47,\ 171m + 14)$.

Now, from the first ascension of the previous paper and from Lemmas 5 and 6, we have the following theorem.

THEOREM 12. *Every positive integer is a sum of B values of* (13) *for* $36 \leqq m \leqq 1950$.

We again follow the procedure of the previous paper, and beginning from an arbitrary point, construct a set of intervals such that

$$B - r - 3 = \left[\frac{m+1}{2}\right] - r$$

values will suffice over each, where we take

$$r = (R - A - \epsilon - 21)/2, \text{ and positive,} \tag{19}$$

and ϵ is 0 or 1 according as $R - A - 1$ is even or odd.

We begin with $f(A+1) = Rm + A + 1$, and our first interval is

$$\big(Rm + A + 1,\ (R+1)m + A - 2r\big).$$

The rest of the intervals are

$$\big((R + t - 1)m + 3R + 3t - 3,\ (R + t)m + R + t - 2r\big), \tag{20}$$

$$\big((R + t)m + R + t - 2r,\ (R + t + 1)m + R + t - 4r - 23\big), \quad (t = 1, \cdots, 10). \tag{21}$$

In constructing the intervals (20) and (21) we take for the initial point of (20) the integer $(R + t - 1)a$. The value for r in (19) was obtained by requiring that r be the greatest integer satisfying

$$T(A, t) = (R + t)m + R + t - 2r \geqq f(A + 1) + ta, \quad (t = 1, \cdots, 10), \tag{22}$$

uniformly in t. Substituting the value for r from (19) in $T(A, t)$ (the end point of (20)), we get

$$T(A, t) = f(A + 1) + ta + (10 - t)2 + \epsilon.$$

Hence $T(A, t)$ is a sum of 12 values of $f(x)$ (of which one value will be zero if $R - A - 1$ is even). From this we obtain the interval (21) in the usual way.

An inspection gives us the information that these intervals will overlap for

$$m \geqq Q(A) = 4R - 2A - 1 = \frac{2}{3}(A^3 - 4A) - 1.$$

Expressing r in terms of A, we have

$$r = \frac{1}{12}(A^3 - 7A) - \frac{\epsilon + 21}{2}.$$

LEMMA 7. *For* $m \geqq Q(A)$, $A \geqq 6$, *every integer* n,

$$Rm + A + 1 \leqq n \leqq (R + 10)m + A + 31,$$

is a sum of $[(m+1)/2] - r$ *values of* $f(x)$.

The analogues to statements (S_1), (S_2), and (S_3) for the polynomial now under consideration are:

(S'_1) For $Q(A) \leqq m < Q(A+1)$ every integer $> f(A+1)$ is a sum of B values, provided $A \geqq 10$.

(S'_2) For $m \geqq Q(A)$, B values will suffice from $f(A+1)$ to $f(A+2)$, when $A \geqq 10$.

(S'_3) For $m \geqq Q(A)$, every integer n, $f(11) \leqq n \leqq f(A+2)$, will be a sum of B values.

As before (S'_3) is obtained from (S'_2), and (S'_1) and (S'_3) are sufficient to prove the following theorem.

THEOREM 13. *Every integer* $\geqq f(11)$ *is a sum of* B *values of* $f(x)$ *for* $m \geqq Q(10) = 639$.

In proving (S'_1) we proceed as we did before. From Lemma 7 and the ascension that established (5), we have that every integer N,

$$f(A+1) \leqq N \leqq c_3 m = \frac{9}{2}(10)^{2(3/2)^r} m, \tag{23}$$

is a sum of B values of $f(x)$ for $m \geqq Q(A)$. It is evident that

$$10^{12} m^{10} \leqq c_3 m, \text{ for } Q(A) \leqq m \leqq M = (10^{-12} c_3)^{1/9}. \tag{24}$$

For $A \geqq 10$ (henceforth we shall require $A \geqq 10$) we have that

$$r > e_1 = \frac{A^3}{16}. \tag{25}$$

From (4_1) we see that

(26) $$\frac{2}{9}\left(\frac{3}{2}\right)^{c_1} - \frac{12}{9} > c_2 = \frac{10^{-4}A^6}{2},$$

and from (4_2) we see that

(27) $$10^{c_2} > 10^{-13}A^{18} > Q(A+1).$$

Therefore $M > Q(A+1)$ from (23), (24), (25), (26) and (27); and we have proved (S'_1). It is evident that

$$c_3 m > Mm > mQ(A) > f(A+2),$$

and therefore by (23) we have established (S'_2).

From Lemmas 5 and 6 and Theorems 12 and 13 we get

THEOREM 14. *Every positive integer is a sum of* $[(m+1)/2]+3$ *values of* $m(x^3-x)/6 - m(x^2-x)/2 + x$ *for non-negative integers* x, *where* $m \geqq 36$.

We embody the results of Theorems 8 and 14 in a final, recapitulative theorem.

THEOREM 15. *Every positive integer is a sum of* $[(m+1)/2]+3$, $m+3$, *or* 9 *values of* $m(x^3-x)/6 - m(x^2-x)/2 + x$ *for non-negative integers* x *according as* $m \geqq 36$, $35 \geqq m \geqq 16$, *or* $6 \geqq m \geqq 1$.

THE UNIVERSITY OF CALIFORNIA,
BERKELEY, CALIFORNIA.

PROOF OF A THEOREM OF LEHMER.[1]

By Paul M. Hummel.

1. Introduction. Galois first proved the theorem that if the partial quotients of a convergent, purely periodic binary continued fraction were inverted (their order reversed), the new continued fraction converged to a number belonging to the same quadratic field as the original. The like is not usually true for continued fractions of higher order as may be seen by taking almost any example at random. Lehmer[2] gave a generalization of Galois' theorem for the ternary continued fraction but remarked that his method of proof could not handle the general, or n-ary, case. The purpose of this article is to give a short elementary proof of Lehmer's theorem for the n-ary continued fraction.

2. Continued fractions of order n. The n-ary continued fraction is defined as follows: let $(x_1', x_1'', \cdots, x_1^{(n-1)})$ be any $n-1$ real numbers; by an n-ary continued fraction representation of this set we shall mean the $(n-1)$-uple sequence of numbers

$$(p_1', p_1'', \cdots, p_1^{(n-1)};\ p_2', p_2'', \cdots, p_2^{(n-1)};\ \cdots;\ p_k', p_k'', \cdots, p_k^{(n-1)}; \cdots)$$

obtained from the recursion formulas

$$\begin{aligned} x_k' &= p_k' + 1/x_{k+1}^{(n-1)}, \\ x_k^{(i)} &= p_k^{(i)} + x_{k+1}^{(i-1)}/x_{k+1}^{(n-1)}, \qquad (i = 2, \cdots, n-1), \end{aligned} \tag{1}$$

where the $p_k^{(i)}$ are chosen by some definite law of selection.

The set of numbers $(p_k', p_k'', \cdots, p_k^{(n-1)})$ is called the k-th *partial quotient set* and is denoted for brevity by p_k. Similarly

$$x_k = (x_k', x_k'', \cdots, x_k^{(n-1)})$$

is called the k-th *complete quotient set.*

In order to apply matric methods, the complete quotients are replaced by homogeneous notation. Thus we set

[1] Received July 23, 1936.

[2] D. N. Lehmer, *Bulletin of the American Mathematical Society* (1931), pp. 565-570.

$$x_k^{(i)} = u_k^{(i+1)}/u_k', \qquad (i = 1, \cdots, n-1).$$

The defining equations (1) are then replaced by the equivalent ones

$$\begin{aligned} u_k' &= u_{k+1}^{(n)}, \\ u_k^{(i+1)} &= u_k^{(i)} + p_k^{(i)} u_{k+1}^{(n)}, \qquad (i = 1, \cdots, n-1). \end{aligned}$$

These equations may be considered as a unimodular transformation whose matrix is

$$T_k = \left\| \begin{matrix} 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & p_k' \\ 0 & 1 & 0 & \cdots & p_k'' \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & p_k^{(n-1)} \end{matrix} \right\|.$$

Since the matrix of a product of transformations is the product of the matrices, it follows that

$$u_1^{(r)} = \sum_{i=1}^{n} A_{k-n+i}^{(r)} u_{k+1}^{(i)}, \qquad (r = 1, \cdots, n),$$

where the matrix

$$M_k = (A_{k-n+s}^{(r)}) = T_1 T_2 \cdots T_k.$$

Since $M_{k+1} = M_k T_{k+1}$, we obtain the new set of recursion formulas

$$A_{k+1}^{(r)} = A_{k-n+1}^{(r)} + A_{k-n+2}^{(r)} p'_{k+1} + \cdots + A_k^{(r)} p_{k+1}^{(n-1)}, \qquad (r = 1, \cdots, n),$$

with initial values, $(A_{-n+s}^{(r)}) = I$.

The set of numbers $(A_k''/A_k', A_k'''/A_k', \cdots, A_k^{(n)}/A_k')$ is by definition, the k-th *convergent set*. In case the $\lim\limits_{k\to\infty} (A_k^{(i)}/A_k')$, $(i = 2, \cdots, n)$, exists, the expansion is said to be *convergent* and it is readily shown[3] that these limits are $x_1^{(i-1)}$ respectively.

In case the partial quotient sets ultimately become periodic, the representation is said to be *periodic*. Let

$$(2) \qquad (p_1; \cdots; \overline{p_{l+1}; \cdots; p_{l+h}})$$

be a periodic n-ary continued fraction of period h with l non-recurring partial quotient sets. We define

$$P = (P_{rs}) = T_{l+1} T_{l+2} \cdots T_{l+h}$$

[3] See O. Perron, *Mathematische Annalen*, vol. 64 (1907), pp. 2-76, or my Dissertation, The Ohio State University (1935).

to be the *period matrix* of the continued fraction. The characteristic equation of P will also be called the characteristic equation of the continued fraction. It is easily shown [4] that if the continued fraction is convergent, the $x_1^{(i)}$ belong to the algebraic field defined by the characteristic equation.

3. Linear continued fractions. If the partial quotients contained in the period of a periodic continued fraction are representable in terms of a linear parameter, the continued fraction is called *linear*.[5] Thus (2) is a linear continued fraction provided

$$p_k^{(i)} = a_i t_k + b_i, \qquad (i = 1, \cdots, n-1;\ k = l+1, \cdots, l+h).$$

An interesting property of this type of continued fraction is given by

LEHMER'S THEOREM.[6] *The characteristic equation of a linear n-ary continued fraction is unaltered by inverting the partial quotient sets contained in the period.*

To prove the theorem, we define an auxiliary matrix A as follows: the first row of A = the first column of $A = (a_1, a_2, \cdots, a_{n-1}, 0)$; the last row of A = the last column of $A = (0, a_1, a_2, \cdots, a_{n-1})$; while the remaining elements of A are determined by the formula

$$(3) \qquad a_{r,s-1} = a_{r-1,s} + D_{s-1,r-1}, \qquad (r = 2, \cdots, n-1;\ s = 3, \cdots, n),$$

where $D_{ij} = (a_i b_j - a_j b_i)$. Thus each row is determined in terms of the preceding row and the D's.

We will first show that the matrix A is symmetric. Interchanging r and s in (3) gives, (since $D_{ij} = -D_{ji}$)

$$(4) \qquad a_{s-1,r} = a_{s,r-1} + D_{s-1,r-1}.$$

The left members of (3) and (4) will be equal if $a_{r-1,s} = a_{s,r-1}$, that is, the r-th row will equal the r-th column provided the $(r-1)$-st row equals the $(r-1)$-st column and since the first row equals the first column, A is symmetric.

We next show that the determinant of A, $|A|$, which is a polynomial in the a_i and the b_i, is not identically zero. This is easily seen by taking all the a_i and b_i except a_1 equal to zero. This gives

[4] See Perron or my Dissertation, *loc. cit.*

[5] So named by Lehmer, *loc. cit.*

[6] Lehmer, *loc. cit.*, proved the theorem for $n = 3$ and 4 but stated that he had not yet found a general proof.

$$A = \left\|\begin{matrix} a_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & a_1 \\ 0 & 0 & 0 & \cdots & a_1 & 0 \\ \vdots & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & a_1 & \cdots & 0 & 0 \\ 0 & a_1 & 0 & \cdots & 0 & 0 \end{matrix}\right\|.$$

Whence $|A| = \pm a_1^n$, which proves the assertion.

Now consider the matric product $S_k = AT_k^T$. ($T^T = T$-transpose). Written at length, this is

$$S_k = \left\|\begin{matrix} a_1 & a_2 & \cdots & a_{n-1} & 0 \\ a_2 & a_{22} & \cdots & a_{2,n-1} & a_1 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n-1} & a_{n-1,2} & \cdots & a_{n-1,n-1} & a_{n-2} \\ 0 & a_1 & \cdots & a_{n-2} & a_{n-1} \end{matrix}\right\| \cdot \left\|\begin{matrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & p_k' & p_k'' & \cdots & p_k^{(n-1)} \end{matrix}\right\|,$$

where, of course, $p_k^{(i)} = a_i t_k + b_i$. We will show that S_k is symmetric. It is apparent that the first row of S_k = the first column of $S_k = (0, a_1, a_2, \cdots, a_{n-1})$. For $(r, s = 2, \cdots, n)$ the elements e_{rs} of S_k are

$$e_{rs} = a_{r,s-1} + a_{r-1}(a_{s-1}t_k + b_{s-1}),$$

and hence

$$e_{rs} - e_{sr} = a_{r,s-1} - a_{s,r-1} + D_{r-1,s-1},$$

which in view of (3) and the fact that A is symmetric, is equal to zero. Therefore S_k is symmetric so that

$$(5) \qquad AT_k^T = T_k A, \qquad (k = l+1, \cdots, l+h).$$

Now let the partial quotients in the period of (2) be inverted. The new period matrix is

$$P^* = T_{l+h}T_{l+h-1} \cdots T_{l+1}.$$

Multiplying on the right by the auxiliary matrix A gives

$$P^*A = T_{l+h} \cdots T_{l+2}(T_{l+1}A) = T_{l+h} \cdots T_{l+3}(T_{l+2}A)T^T_{l+1},$$

and by repeated use of (5), we obtain $P^*A = AP^T$. Whence

$$(P^* - \lambda I)A = A(P^T - \lambda I),$$

where λ is an indeterminate. Taking determinants gives

$$|P^* - \lambda I| \cdot |A| = |A| \cdot |P^T - \lambda I|.$$

Since $|A|$ is a polynomial in the a_i and b_i which is not identically zero, it may be divided out, giving

$$|P^* - \lambda I| = |P^T - \lambda I|,$$

and this relation holds identically when the a_i, b_i, and λ are considered as independent indeterminates. Hence the characteristic function of P^T, and therefore P, is the same as that of P^*. This proves the theorem.

4. Conclusions. There are two other obvious types of continued fractions for which the characteristic equation is invariant under inversion of the partial quotients. First, those in which the partial quotient sets read the same backward and forward. Second, those which have just two sets of partial quotients in the period. This latter statement is true because of the well known matric theorem that the characteristic equation of a product of matrices is invariant under any cyclic permutation of the matrices. It also follows indirectly from Lehmer's theorem.

Whether there are types other than the three listed for which the characteristic equation is invariant under inversion is not known.

THE UNIVERSITY OF ALABAMA.

REPRESENTATION OF LARGE INTEGERS BY CUBIC POLYNOMIALS.[1]

By MARY HABERZETLE.

By use of the prime number theory we are able to prove the following theorem regarding the representation of all sufficiently large integers. Complete proof of this theorem and the lemmas is to be found in my dissertation at the University of Chicago. Only the theorem is proved here.

Consider the ternary forms:

$$A = x_1^2 + x_2^2 + x_3^2, \quad B = x_1^2 + x_2^2 + 2x_3^2, \quad C = x_1^2 + x_2^2 + 5x_3^2,$$
$$D = x_1^2 + 2x_2^2 + 2x_3^2, \quad E = x_1^2 + 2x_2^2 + 3x_3^2, \quad F = x_1^2 + 2x_2^2 + 4x_3^2,$$
$$G = x_1^2 + 2x_2^2 + 6x_3^2, \quad H = x_1^2 + 2x_2^2 + 5x_3^2.$$

Denote a general one of these by $f = h_1x_1^2 + h_2x_2^2 + h_3x_3^2$.

THEOREM. *Let $m = 1, 2, 3, 4,$ or 5, and let t be a given positive integer prime to 6 or 30. For t prime to 6, let h_1, h_2, h_3 be given by any one of the forms, $A, B, \cdots, G$, and for t prime to 30, let h_1, h_2, h_3 be given by any one of the forms, $A, B, \cdots, H$. Then all sufficiently large integers are represented by*

$$n = ma^3 + tb^3 + \sum_{i=1}^{3} h_i(c_i^3 + d_i^3).$$

We make use of two lemmas.

LEMMA 1. *If p is a prime $\equiv 2 \pmod 3$, and if m is any integer not divisible by p, every integer not divisible by p is congruent modulo p^n to the product of a cube by m.*

A proof of Lemma 1 has been given by L. E. Dickson.[2]

LEMMA 2. *If t is a positive integer prime to 6, every integer $\geqq 23^3t$ is represented by $t\gamma^3 + 6f$, where $\gamma \geqq 0$ and $f = A, \cdots, G$. If t is a positive integer prime to 30, every integer $\geqq 23^3t$ is represented by $t\gamma^3 + 6f$, where $\gamma \geqq 0$ and $f = A, \cdots, H$.*

---

[1] Received August 10, 1936; revised October 29, 1936.

[2] L. E. Dickson, "Simpler proofs of Waring's theorem on cubes, with various generalizations," *Transactions of the American Mathematical Society*, vol. 30 (1928), p. 2.

Proof of Theorem. Let $w = 2(h_1 + h_2 + h_3)$, and let r be the real ninth root of $(w + m + \xi t)/(w + m + \eta t)$, where ξ is a positive number such that $m + \xi t \leqq 6$, and where $0 < \eta < \xi$. Then $r > 1$. It is known that the number of primes $\equiv 2 \pmod 3$ which exceed x and are $\leqq rx$ increases indefinitely with x. Choose as x the first radical in (1). Then for all sufficiently large integers n, there exist at least ten primes p such that

$$(1) \quad [n/(w + m + \xi t)]^{1/9} < p \leqq [n/(w + m + \eta t)]^{1/9}, \quad p \equiv 2 \pmod 3.$$

The product of the ten primes exceeds $[n/(w + m + \xi t)]^{10/9}$ and hence exceeds n if $n > (w + m + \xi t)^{10}$. Hence, not all ten are divisors of n. In what follows, let p be a prime $> m$, not dividing n, and satisfying (1). By Lemma 1, there exist integers a and M satisfying

$$n \equiv ma^3 \pmod{p^3}, \qquad n - ma^3 = p^3 M, \qquad 0 < a < p^3.$$

By (1),

$$(w + m + \eta t)p^9 \leqq n < (w + m + \xi t)p^9.$$

Therefore,

$$(w + m + \eta t)p^9 - mp^9 < n - ma^3 = p^3 M,$$
$$p^3 M < n < (w + m + \xi t)p^9.$$

Cancellation of the factor p^3 gives

$$(w + \eta t)p^6 < M < (w + m + \xi t)p^6.$$

Write $M = N + wp^6$. Then

$$\eta t p^6 < N < (m + \xi t)p^6 \leqq 6p^6.$$

For n sufficiently large, $\eta p^6 \geqq 23^3$. Then $N > 23^3 t$. By Lemma 2, N can be represented by $t\gamma^3 + 6f$, $f = h_1 x_1^2 + h_2 x_2^2 + h_3 x_3^2$, $\gamma \geqq 0$. Then

$$\begin{aligned} n &= ma^3 + t(p\gamma)^3 + 6fp^3 + wp^9 \\ &= ma^3 + t(p\gamma)^3 + \sum_{i=1}^{3} h_i[(p^3 + x_i)^3 + (p^3 - x_i)^3]. \end{aligned}$$

If any $|x_i| \geqq p^3$, $N \geqq 6p^6$, contrary to the above that $N < 6p^6$. Therefore, each cube is $\geqq 0$. This proves the theorem.

THE UNIVERSITY OF CHICAGO.

GROUPS OF ORDER 64 WHOSE SQUARES GENERATE THE FOUR GROUP.[1]

By G. A. Miller.

1. *General theorems.* When the squares of the operators of a group generate the four group its order is obviously of the form 2^m and the number of these groups whose commutator subgroup is of order 2 is known to be $7m/2 - 12$ when $m > 2$ is even and $7(m-1)/2 - 9$ when $m > 3$ is odd. The order of the commutator subgroup of such a group cannot exceed 4 since the quotient group with respect to the subgroup generated by the squares is abelian. When the order of this commutator subgroup is 4 it is the four group and it appears in the central of the group. The determination of all the groups which have the four group both for the group of their squares and also for their commutator subgroup presents many difficulties which remain unsolved. The order of such a group is at least 32 and it is known that there are 9 such groups of this order.[2] The present article is devoted to the determination of some general properties of these groups by means of which all such groups of order 64 can be readily determined. These properties seem to exhibit a better method for the determination of all the groups of order 64 than the one used in an earlier volume of the present journal.[3]

Let G represent a group which has the four group both for the group of its squares and also for its commutator subgroup. When the order of G is 32 then it contains an abelian subgroup of order 16 and each of its other operators has four conjugates. Hence none of these operators appears in an abelian subgroup of order 16. Whenever the order of G exceeds 32 then every one of its operators appears in such a subgroup. This theorem applies to all the groups which have the four group as the group generated by their squares since this four group is then in the central of the group and the corresponding quotient group is the abelian group of type 1^n. To prove the theorem in question let s represent any operator of the group and consider the subgroup composed of the operators which are commutative with s. This subgroup is at least of order 16. If it is of this order it includes the central of G and

[1] Received July 18, 1936.

[2] G. A. Miller, *Proceedings of the National Academy of Sciences*, vol. 22 (1936), pp. 112-115.

[3] G. A. Miller, vol. 52 (1930), pp. 617-636.

hence it is abelian. If it is of larger order it includes an abelian subgroup of order 16 which contains s.

When G is of order 64 all of its operators appear in five subgroups of order 16 which have the group of the squares in common and satisfy the condition that no two of them have any other common operator. This results from the fact that the abelian group of order 16 and of type 1^4 has the property that all of its operators appear in five subgroups and hence every two of these subgroups have only the identity in common. When the order of G exceeds 64 it contains a subgroup of order 64 which satisfies the given conditions. The first of these five subgroups can be chosen arbitrarily and the second can then be any of those which has only the group of the squares in common with the first. To prove that these five subgroups can be so chosen that at least two of them are abelian it is only necessary to prove that we can use for the second an abelian group after the first has been so chosen that it is abelian. When the first of these is contained in an abelian subgroup of index 2 this is obvious.

When the first of these subgroups is not contained in an abelian subgroup of index 2 we may consider the abelian subgroup of order 16 which includes an operator not contained in the first of these subgroups. If this second abelian subgroup has only four operators in common with the first no further proof is necessary. If it has eight operators in common with it the two generate a group of order 32 whose central is of order 8 and which involves exactly three abelian subgroups of order 16. The abelian subgroup of order 16 which contains an operator which is not found in this subgroup of order 32 has 8 operators in common with this subgroup of order 32. These common operators may be assumed to include the group of the squares of G. If they do not constitute the central of this subgroup of order 32 the first abelian subgroup of order 16 can clearly be so chosen as to have only four operators in common with the second.

In the other possible case G involves an abelian subgroup of index 2 since its central is of order 8, in view of the theorem which will be proved in § 3 that a necessary and sufficient condition that a group of order 64 whose commutator subgroup is the four group contains an abelian subgroup of index 2 is that its central is of order 8. This proves the following theorem: *Every group of order 2^m, $m > 5$, which has the four group for the group of its squares contains two abelian subgroups of order* 16 *which have only this four group in common.*

2. *Groups of order* 64 *which contain only five abelian subgroups of order* 16. Unless the contrary is stated it will always be assumed in what follows

that both the commutator subgroup and the group of the squares of G is the four group. When G is of order 64 it is generated by the two abelian subgroups of order 16 noted at the close of the preceding paragraph. Suppose that every operator of G which is not contained in the subgroup generated by its squares has four conjugates and hence that no two abelian subgroups of order 16 have more than four common operators. In this case G contains five and only five abelian subgroups of order 16 and these subgroups include all the operators of G and every two of them have only the group of the squares in common. It should be noted that G contains also other sets of five subgroups of order 16 such that they involve all the operators of G and that no two of them have any operators in common besides those which appear in the commutator subgroup but each of these sets includes non-abelian subgroups of order 16.

There is obviously one and only one such G in which the two generating abelian subgroups are both of type 1^4. This contains 36 operators of order 4 and each of its three other abelian subgroups of order 16 is of type 2^2. There is also one and only one such G which is generated by two abelian subgroups of types 1^4 and $2, 1^2$ respectively. It also contains 36 operators of order 4 and two of its other abelian subgroups of order 16 are of type $2, 1^2$ while the remaining one is of type 2^2. When one of the generating abelian subgroups is of type 1^4 it is not possible that each of the others is of type 2^2 since some of the remaining operators are clearly of order 2. Hence there are two and only two such groups of order 64 in which a pair of generating abelian subgroups of order 16 includes the group of type 1^4.

Suppose now that G contains no abelian subgroup of type 1^4 but that at least one of its abelian subgroups is of type $2, 1^2$. Since not all of its remaining operators could be of order 4 it results that such a G contains at least two abelian subgroups of type $2, 1^2$ and hence it is generated by two such subgroups. When the operators of order 4 in these two subgroups have a common square and operators of order 2 contained therein transform each other into themselves multiplied by this square there results a G which involves 52 operators of order 4. Its other three abelian subgroups of order 16 are therefore of type 2^2. When they transform each other into themselves multiplied by a different square there results a G which involves 44 operators of order 4. Two of its other three abelian subgroups of order 16 are of type $2, 1^2$ and their operators of order 4 have the same square but this square is not the same as the square of the operators of order 4 in the two given generating abelian subgroups of order 16.

When G contains 60 operators of order 4 each of its five abelian subgroups

of order 16 is of type 2^2. An operator of one of two such generating subgroups could not transform all the operators of the other into their inverses since G would then contain more than three operators of order 2. For the same reason an operator of one of these subgroups must transform at least one operator of order 4 in the other into its inverse. It must therefore transform four of these operators into their inverses and the other two sets of four into themselves multiplied by the common square of the other set of four. Hence the transformation is completely determined and there is one and only one such group of order 64. Hence *there are five and only five groups of order* 64 *which separately satisfy the conditions that both their commutator subgroup and the group generated by their squares is the four group and that each of their other operators has four conjugates.*

3. *Groups which contain an abelian subgroup of index* 2. When G is the direct product of a group of order 32 which has the four group both for the group of its squares and for its commutator subgroup it contains an abelian subgroup of index 2 but the 9 groups which satisfy this condition will be excluded from the considerations of the present section. Hence we shall not consider here the case when this abelian subgroup is of type 1^5. It was noted above that a necessary and sufficient condition that G contains an abelian subgroup of index 2 is that its central is of order 8. The necessity of this condition results directly from the fact that the order of the commutator subgroup of G is 4. Moreover, when the central of G is 8 and it contains a non-abelian subgroup of index 2 this subgroup has a commutator subgroup of order 2 and hence some of the operators of G which are not found in this subgroup are commutative with 16 of its operators. In all the other groups considered in the present article the central is therefore identical with the commutator subgroup.

The abelian subgroup of index 2 contained in G is characteristic since it is the only abelian subgroup of index 2 contained in G. When it is of type $2, 1^3$ then its central is of type $2, 1$ and hence there is one and only one such group. It contains 40 operators of order 4. When the abelian subgroup of index 2 is of type $2^2, 1$ the central of G is again of type $2, 1$ and two co-sets with respect to it are similar and involve only operators of order 4. The third co-set involves operators which have the same square as those of the central. There are two groups when the operators of this co-set are transformed into themselves multiplied by the square of an operator of order 4 in the central. One of these contains 40 operators of order 4 and 24 of these have a common square while the other contains 56 operators of order 4 composed of 8, 24, 24

operators respectively which have a common square. A third group results when the operators of the said co-set are transformed into themselves multiplied by an operator of order 2 which is not the square of an operator of order 4 in the central. This involves 48 operators of order 4. Hence, excluding direct products, *there are four groups of order* 64 *which involve an abelian subgroup of index* 2 *and have the four group both for their commutator subgroup and also for the group of their squares.*

4. *Some properties of the remaining groups.* In each of the remaining groups every pair of generating abelian subgroups of order 16 which have only the central of G in common must satisfy the condition that at least four operators of each of these subgroups have two and only two conjugates under the other since not every operator which is not in the central of G can have four conjugates. If one of these two subgroups has only four such operators the other will also contain only four such operators and the operators of G which have two conjugates will generate an abelian subgroup of order 16 involving 12 such operators. Each of these two generating abelian subgroups will then contain 8 operators which have four conjugates under the other and G involves exactly seven abelian subgroups of order 16. In the second possible case each of these two generating abelian subgroups contains 8 operators which have only four conjugates under the other. As four of these must give rise to a commutator which differs from the commutator to which the other four give rise each of these generating subgroup then contains 4 and only 4 operators which have eight conjugates under the other. The operators which have only two conjugates under G in this case generate G and such a G contains exactly nine abelian subgroups of order 16.

In order to simplify the considerations which follow it is desirable to note that with the exception of the groups which involve 60 operators of order 4 each of the remaining groups contains a pair of generating abelian subgroups of order 16 such that at least one of them is of type $2, 1^2$. To prove this theorem it is only necessary to prove that such a G contains at least one abelian subgroup of type $2, 1^2$ which involves the commutator subgroup. For if another abelian subgroup of order 16 would have eight operators in common with this one the two would generate a group of order 32 whose commutator subgroup would be of order 2. An abelian subgroup of order 16 containing an operator not found in this subgroup of order 32 would therefore have only the commutators of G in common with the given group of type $2, 1^2$. If G did not contain such a group its operators of order 2 which are not found in its central would appear in one or more subgroups of type 1^4. Hence it is only

necessary to prove that if G contains a subgroup of this type it must also contain one of type $2, 1^2$.

If G contains two subgroups of type 1^4 which have only the central of G in common then each of its four abelian subgroups of order 16 which has only the central in common with one of them contains four different operators from one of the three co-sets of G with respect to this subgroup, viz., from one which involves 12 operators of order 4. It therefore results that G involves an abelian subgroup of type $2, 1^2$. If it contains only one subgroup of type 1^4 then not all of the remaining operators can be of order 4 and hence there is again a subgroup of type $2, 1^2$. It therefore follows that *every group of order* 64 *which does not involve* 60 *operators of order* 4 *but has the four group for its commutator subgroup and also for the group of its squares contains the abelian subgroup of type* $2, 1^2$ *unless each of its operators which is not in its central has four distinct conjugates.*

5. *Groups in which the operators which have two conjugates generate an abelian subgroup of order* 16. The operators of two of the co-sets of G with respect to one of a pair of generating abelian subgroups have four conjugates under this subgroup while the operators of the third of these co-sets have only two conjugates under the same subgroup but half of these operators have four conjugates under G. The four abelian groups of order 16 which have only the central in common with the given group of this order have eight operators in common in pairs. Four of these belong to one of the co-sets with respect to the given abelian subgroup of order 16 but their operators belonging to the other two of these co-sets are distinct. If one of the two generating abelian subgroups of order 16 is of type $2, 1^2$ and the second is of type 1^4 the resulting G contains 32 operators of order 4 and the three additional abelian subgroups which have only the central in common with the first are of type $2, 1^2$. This G is also generated by two abelian subgroups of type 1^4 and hence there is only one such group.

When G involves an abelian subgroup of type 1^4 but no such subgroup which has only the central in common with it, it contains 48 operators of order 4 and these have only two distinct squares in sets of 24. Hence there are two and only two groups which come under the present heading and in which the operators having two conjugates generate the abelian group of type 1^4. When the operators which have two conjugates generate the abelian group of type $2, 1^2$ and G involves a second abelian subgroup of this type which has 8 operators in common with it these 8 operators may constitute the group of type 1^3 or the group of type $2, 1$. In the former case there are two groups

which involve separately 48 operators of order 4. In one of these 32 operators of order 4 have the same square while in the other at most 24 operators have this property. When the 8 operators of the abelian group of order 16 and of type $2, 1^2$ which is generated by the operators which have two conjugates and appear in the given abelian subgroup of type $2, 1^2$ constitute the group of type $2, 1$ it may be assumed that at least one of the four generating abelian subgroups of order 16 which have only the central in common with the given one of type $2, 1^2$ is either of type 1^4 or of type $2, 1^2$.

When the first of these two conditions is satisfied there are two groups involving 32 and 40 operators of order 4 respectively. When this condition is not satisfied none of the remaining four generating abelian subgroup in question is of type 1^4 but at least one of them is of type $2, 1^2$. There are three such groups involving 40, 48, 48 operators of order 4 respectively. Only one of the last two involves 32 operators which have a common square. Hence there are 7 such distinct groups of order 64 which satisfy the condition that the operators which have two conjugates generate the abelian group of type $2, 1^2$, and 2 in which these operators generate the abelian group of type 1^4. It remains to determine those in which these operators generate the abelian group of type 2^2 and hence the given subgroup of type $2, 1^2$ involves four operators of order 4 which have two conjugates under the group, and all the operators which have two conjugates are of this order. There are three such groups involving 40, 48, 56 operators of order 4 respectively. Since a group in which the operators which have two conjugates generate an abelian group of order 16 could not contain 60 operators of order 4 *there are* 12 *and only* 12 *groups of order* 64 *which have the four group both for their commutator subgroup and for the group of their squares and satisfy the condition that their operators which have two conjugates generate an abelian subgroup of order* 16.

6. *Groups which are generated by their operators which have two conjugates.* Since the four group is assumed to be the central of each of these groups each of a pair of generating abelian subgroups of order 16 involves 8 operators which have two and only two conjugates under the other. There are 8 other operators in G which have two and only two conjugates while the remaining 36 operators of G have four conjugates. For each of the 9 abelian subgroups of order 16 contained in G there are 4 others which have four operators in common with it and 4 others which have eight such common operators. We shall first determine all these groups which involve at least one abelian subgroup of type 1^4.

There is one and only one group in which a pair of generating abelian subgroups of order 16 is composed of groups of type 1^4. It contains 28 operators of order 4. Two sets of 12 of these operators have a common square but the squares of one set are not the same as those of the other. The remaining 4 have the third commutator of order 2 for their common square. When G contains at least one abelian subgroup of type 1^4 there is a pair of generating abelian subgroups of order 16 which includes it. We proceed to determine those groups which involve one such subgroup but not a pair of such generating subgroups. Two of the co-sets of the group with respect to this subgroup of type 1^4 are conjugate. If we multiply the operators of these co-sets successively by operators of order 4 whose squares appear therein there result two conjugate groups of order 64 involving 36 operators of order 4 but if we multiply the operators of these co-sets by an operator of order 4 having a different square there results a group of order 64 which has 44 operators of this order. Hence *there are three such groups of order* 64 *which involve separately* 9 *abelian subgroups of order* 16 *and include such a subgroup of type* 1^4.

It remains to determine the groups which do not contain an abelian subgroup of type 1^4. Near the close of § 2 it was noted that there is one and only one group of order 64 which contains 60 operators of order 4 and in which every operator of this order has four conjugates. Each of the other groups determined above involves less than 60 operators of order 4. Any additional groups which belong to the category of groups considered in the present article and involve 60 operators of order 4 must therefore come under the present section. We proceed to determine these groups.

A pair of abelian generating subgroups of order 16 in such a group which have only the central in common must satisfy the conditions that each of them is of type 2^2 and that each involves two independent generators such that one of them is commutative with one of the other having a different square and transforms the second of the other into its inverse. As the co-set which involves the former of these generators involves operators of order 4 having two distinct squares the group of order 32 thus obtained is completely determined. A second independent generator of the former of these subgroups can be chosen in two essentially different ways and hence there are two such groups which involve 60 operators of order 4. In one of these 36 operators of order 4 have the same square while in the other at most 28 of these operators satisfy this condition. Hence *there are three and only three groups of order* 64 *which satisfy the conditions that they separately involve* 60 *operators of order* 4 *and have the four group both for the group of their squares and for their commutator subgroup.* Only two of these three groups belong to the present section.

Since the remaining groups which belong to the present section cannot involve an abelian subgroup of type 1^4 and cannot involve 60 operators of order 4 they must involve an abelian subgroup of type $2, 1^2$ and the four other abelian subgroups which have only the central in common with this subgroup are of one or more of the following two types: $2, 1^2$; 2^2. Suppose first that G is generated by two such subgroups of type $2, 1^2$ whose operators of order 4 have the same square. Not all of their operators of order 2 can be commutative since these groups are supposed to contain no abelian subgroup of type 1^4. There are two such groups in which all the operators of one of these subgroups are transformed into their inverses by an operator of the other. One of these contains 44 operators of order 4 while the other contains 52 such operators. Each of these groups contains four other abelian subgroups of type $2, 1^2$ and the operators of order 4 in these have the same square as in the two given generating subgroups of this type.

Suppose that G is generated by two abelian subgroups of type $2, 1^2$ whose operators of order 4 have the same squares and that one of these subgroups again contains an operator which is commutative with the operators of order 2 in the other but which does not transform into their inverses the operators of order 4 in this other. There are three such groups. Two of these involve 44 operators of order 4 while the third involves 52 such operators. The last of these groups can easily be distinguished from the one noted in the preceding paragraph which contains the same number of operators of order 4 by the fact that the latter contains 36 such operators having a common square while the former contains at most 28 such operators. To distinguish the three given groups involving 44 operators of order 4 it may be noted that in one of these there are at most 20 operators of order 4 which have a common square while in each of the others there are 28 such operators. The latter can easily be distinguished by the fact that in one of them all the operators of order 4 in the abelian subgroups of type $2, 1^2$ have a common square while in the other these operators have two different squares.

We have now determined the five groups which are generated by two abelian subgroups of type $2, 1^2$ such that one of these subgroups contains operators which are commutative with the operators of order 2 in the other but not with its operators of order 4 and that the operators of order 4 in these generating subgroups have a common square. If the two generating abelian subgroups of order 16 satisfy these conditions with the exception that operators of order 4 of the one are commutative with operators of order 4 of the other there result two additional groups. One of these can readily be distinguished from those which precede by the fact that it involves 36 opera-

tors of order 4. The second involves 44 such operators but only four abelian subgroups of type $2, 1^2$ while the preceding group which contains 44 operators of order 4 involve a larger number of such subgroups. Hence *there are seven and only seven groups of order* 64 *which separately satisfy the three conditions that both the group of their squares and their commutator subgroup is the four group, that each of them is generated by its operators which have two conjugates and also by two abelian subgroups of type* $2, 1^2$ *whose operators of order* 4 *have a common square.*

If a group is generated by two abelian subgroups of type $2, 1^2$ whose squares are different but not by two such subgroups whose squares are the same it may be assumed that these two squares are fixed. There is one such group in which all the operators of order 2 in one of these generating subgroups are invariant under 8 operators of the other. This group contains 52 operators of order 4 and four abelian subgroups of type $2, 1^2$ but those whose operators of order 4 have the same square have more than 4 common operators. There is also one such group in which there are operators of order 4 in one of these subgroups which are commutative with 8 operators of the other. This group also contains 52 operators of order 4 and four abelian subgroups of type $2, 1^2$. These appear in two sets of two each such that these operators of order 4 have a common square but each such pair has 8 operators in common.

It remains to consider the groups which separately involve no two generating abelian subgroups of type $2, 1^2$ but involve at least one subgroup of this type. There is one such group. It contains 52 operators of order 4 and two abelian subgroups of type $2, 1^2$ but these have eight operators in common and hence they do not generate G. Each of its remaining seven abelian subgroups of order 16 contained therein is of type 2^2.

UNIVERSITY OF ILLINOIS.

NECESSARY AND SUFFICIENT CONDITIONS FOR GENERATING CERTAIN SIMPLE GROUPS BY TWO OPERATORS OF PERIODS TWO AND THREE.[1]

By Abraham Sinkov.

1. *Introduction.* Necessary and sufficient conditions for generating the simple groups of order 60, 168, and 660 by two operators of periods two and three are already known.[2] Professsor G. A. Miller[3] has shown that no such definition is possible for the simple group of order 360. The purpose of the present study is to determine necessary and sufficient conditions for the two remaining simple groups G_{504} and G_{1092}, whose orders do not exceed 1092.

In addition to the above-mentioned groups, G_{5616} has been studied by K. E. Bisshopp[4] who showed that there are only two abstractly distinct ways in which this group may be generated by two operators of periods two and three. However, each of his abstract definitions required nine distinct restrictions upon the generators to insure sufficiency and apparently no attempt was made to study their independence. It will be shown in the present study that five conditions are sufficient to define the group completely in either case.

2. *General relations.* The various definitions which will be obtained in this study will all be based on the relations $S^3 = T^2 = (ST)^n = (S^{-1}T^{-1}ST)^p = 1$, which we shall designate $(2, 3, n; p)$. However, instead of employing the generators S and T, we shall introduce the following substitution:[5]

$$P = (ST)^{-1} \qquad S = P^2Q$$
$$Q = (ST)^2S \qquad T = P^3Q.$$

[1] Received April 22, 1936.

[2] H. R. Brahana, "Pairs of generators for the known simple groups whose orders are less than one million," *Annals of Mathematics*, vol. 31 (1930), pp. 543-544. This paper will hereafter be referred to as "Pairs of generators."

[3] G. A. Miller, *Bulletin of the American Mathematical Society*, vol. 7 (1900-1901), p. 426.

[4] *Bulletin of the American Mathematical Society*, vol. 37 (1931), p. 99.

[5] This same substitution for $(ST)^7 = 1$ was used by Professor Brahana in his paper "Certain perfect groups generated by two operators of periods two and three," *American Journal of Mathematics*, vol. 50 (1928), p. 349. This paper will hereafter be referred to as "Certain perfect groups."

Since $Q = S^{-1}(S^{-1}TST)S$, its period is the same as that of the commutator of S and T. Hence the relations $(2, 3, n; p)$ will be replaced by

$$P^n = Q^p = (QP^3)^2 = (QP^2)^3 = 1.$$

The set of relations $(2, m, n; p)$, i. e.

$$S^m = T^2 = (ST)^n = (S^{-1}T^{-1}ST)^p = 1$$

is intimately connected with the relations

$$A^m = B^n = C^p = (AB)^2 = (BC)^2 = (CA)^2 = (ABC)^2 = 1$$

which have been studied by Dr. H. S. M. Coxeter.[6] He has designated the group defined by these relations $G^{m,n,p}$ and has shown that if any one of the numbers m, n, p is even (say p) then $G^{m,n,p}$ contains as a subgroup of index two the group defined by $(2, m, n; p/2)$. Should two of the numbers be odd (say n and p) then, by putting $Q = C^2$; $P = A$, he is enabled to express the relations $G^{m,n,p}$ in the form

$$\begin{aligned} P^n = Q^p = (Q^{(p+1)/2}P)^2 = (QP^2)^m \\ = \{(QP^2)^{(m-1)/2}P\}^2 = \{Q^{(p+1)/2}(QP^2)^{(m-1)/2}\}^2 = 1. \end{aligned}$$

These relations are not all independent; in fact any one of the last three may be considered redundant. Putting $m = 3$, we may write

$$G^{3,n,p} \equiv (2, 3, n; p); \qquad (Q^{(p+1)/2}P)^2 = 1$$

and observe that either one of the relations $(QP^3)^2 = 1$, $(QP^2)^3 = 1$ may be replaced by $(Q^{(p+3)/2}P^2)^2 = 1$.

For convenience in some of the manipulation which will be required throughout this paper, we will now establish some consequences of the relations $(2, 3, n; p)$.

$$QP^2Q = P^{-2}Q^{-1}P^{-2} = P \cdot P^{-3}Q^{-1} \cdot P^{-2}$$

hence

$$QP^2Q = PQP \tag{1}$$

$$\begin{aligned} P^2QPQ^a &= P \cdot PQP \cdot Q^a = PQP^2Q^{a+1} \\ &= Q \cdot P^2QPQ^a \cdot Q = Q^{-a} \cdot P^2QPQ^a \cdot Q^{-a} \\ &= Q^{-a}P^{-1} \cdot P^3Q \cdot P = Q^{-a}P^{-1}Q^{-1}P^{-2} \end{aligned}$$

whence

$$(P^2QPQ^a)^2 = 1 \quad \text{and also} \quad (PQP^2Q^{a+1})^2 = 1. \tag{2}$$

---

[6] These results, which have been communicated to me by Dr. Coxeter, have not yet been submitted for publication.

From (1)

$$QP^{-1} = P^{-2}Q^{-1}PQ = P \cdot P^{-3}Q^{-1} \cdot PQ$$
$$= PQP^4Q \tag{3}$$
$$P^3Q \cdot P^{-1} = P^4QP^4Q = Q^{-1}P^{-4}$$
$$(QP^4)^3 = 1. \tag{4}$$

From (2)

$$P^2Q^2 \cdot PQP \cdot PQ^2PQ = 1$$
$$P^2Q^2P^2Q \cdot PQP \cdot Q = 1$$
$$(Q^2P^2)^3 = 1. \tag{5}$$

These results lead to some new properties in connection with the independence of the relations defining $G^{3,n,p}$ (n, p odd). We have from (2)

$$QPQ^{(p-1)/2}\, P^2 = P^{-2}Q^{-[(p-1)/2]}\, P^{-1}Q^{-1} = P^{-2}Q \cdot Q^{-[(p+1)/2]}P^{-1} \cdot Q^{-1}.$$

If we assume $(Q^{(p+1)/2}P)^2 = 1$, the right member becomes $P^{-2}QPQ^{(p-1)/2}$. Hence $P^2 \cdot QPQ^{(p-1)/2} \cdot P^2 = QPQ^{(p-1)/2}$. If now, the period of P is an odd number, this leads to

$$QPQ^{(p-1)/2} = PQPQ^{(p-1)/2}P = PQP \cdot Q^{(p-1)/2} \cdot PQ^{(p+1)/2} \cdot Q^{-[(p+1)/2]}$$
$$= QP^2Q \cdot Q^{(p-1)/2}Q^{-[(p+1)/2]}P^{-1}Q^{-[(p+1)/2]} = QPQ^{-[(p+1)/2]}$$

whence $Q^p = 1$.

Therefore, the relations $(2, 3, n)$; $(Q^{(p+1)/2}P)^2 = 1$ imply $Q^p = 1$ and give a complete definition of $G^{3,n,p}$ provided n and p are both odd. These relations are independent whenever $n > 5$, $p > 5$.

In the same way, starting with the operator $QPQ^{(p+3)/2}P^2$, it can be shown that the relations $(2, 3, n; p)$; $(Q^{(p+3)/2}P^2)^2 = 1$ imply $(Q^{(p+1)/2}P)^2 = 1$.

3. *The simple group of order* 504. We proceed first to determine the number of different ways in which G_{504} may be generated by two operators of periods two and three.[7] The 63 operators of period two are all conjugate [8] and it is therefore sufficient to consider only one of them, say

$$T \equiv (1, 3)(2, 6)(4, 5)(7, 8)$$

as a possible generator. The largest subgroup of G within which T is invariant is abelian, of order eight and type $(1, 1, 1)$. It is generated by T and the two operators

[7] The general procedure to be followed has been outlined by Professor Brahana in "Pairs of generators," pp. 542, 543.

[8] The properties of G_{504} and G_{1092} which are made use of here as well as the method of representing these groups as permutation groups are given in Dickson's *Linear Groups with an Exposition of the Galois Field Theory* (1901), chap. XII.

$$T' \equiv (1,4)(2,8)(3,5)(6,7); \quad T'' \equiv (1,7)(2,5)(3,8)(4,6).$$

Under this subgroup, the 56 operators of period three are divided up into seven sets of conjugate operators; the members of each set satisfy with T the same abstract relations. We give below one operator from each set and note the period of its product with T.

	Period of ST
$S_1 = (2,5,3)(4,7,6)(1,9,8)$	9
$S_2 = (3,6,4)(5,1,7)(2,9,8)$	9
$S_3 = (4,7,5)(6,2,1)(3,9,8)$	7
$S_4 = (5,1,6)(7,3,2)(4,9,8)$	9
$S_5 = (6,2,7)(1,4,3)(5,9,8)$	7
$S_6 = (7,3,1)(2,5,4)(6,9,8)$	7
$S_7 = (1,4,2)(3,6,5)(7,9,8)$	2

The permutation $D \equiv (1,2,4)(3,6,5)$ which corresponds to the transformation $Z' = Z^2$ transforms G_{504} into itself according to an outer automorphism. It is commutative with T and transforms S_1 into S_2, S_2 into S_4. Hence the relations satisfied by T with S_1 are the same as those satisfied by T with either S_2 or S_4. Similarly D transforms S_3 into S_6, S_6 into S_5. Hence these three operators satisfy the same relations with T. Since T and S_7 generate a dihedral group of order 6, it follows that there are only two posssible definitions for G_{504}.

Now, it has already been shown by Professor Brahana [9] that the relations $(2,3,7;9)$; $(Q^6P^2)^2 = 1$ are sufficient to define G_{504}. It follows from the last result in Art. 2 that these relations imply $(Q^5P)^2 = 1$. Hence $G_{504} \equiv G^{3,7,9}$. This result enables us to set down at once a complete definition for both cases.

THEOREM. *Two operators of periods two and three generate G_{504} if and only if they satisfy one of the following sets of independent relations*

$$A: \quad (2,3,7); \; (Q^5P)^2 = 1$$
$$B: \quad (2,3,9); \; (Q^4P)^2 = 1.$$

It can be shown that the condition $(Q^5P)^2 = 1$ in A may be replaced by $(Q^6P^2)^2 = 1$. Hence the relations used by Professor Brahana to define G_{504} are not independent.

4. *The simple group of order* 1092. The 91 operators of period two con-

[9] "Certain perfect groups," p. 354.

tained in G_{1092} are all conjugate and it is again sufficient to consider only one of them. We choose it to be

$$T \equiv (1, 12)(2, 11)(3, 10)(4, 9)(5, 8)(6, 7).$$

The largest subgroup of G within which T is invariant is the dihedral group of order 12 generated by

$$(1, 4, 3, 12, 9, 10)(2, 8, 6, 11, 5, 7)$$

and

$$(1, 12)(2, 6)(3, 4)(7, 11)(9, 10)(13, 14).$$

Under this subgroup the 180 remaining operators of period three are divided up into 16 conjugate sets. Of these, two contain only six distinct operators each; the remaining 14 sets each contain 12 operators. As in the case of G_{504}, we give below one member of each of these sets together with the period of its product with T.

	Period of ST
$S_1 = (1, 13, 10)(2, 3, 6)(4, 9, 11)(5, 12, 7)$	6
$S_2 = S_1^2$	6
$S_3 = (2, 10, 4)(11, 13, 5)(3, 6, 7)(8, 12, 9)$	6
$S_4 = S_3^2$	6
$S_5 = (3, 10, 12)(4, 6, 13)(5, 11, 8)(14, 9, 7)$	3
$S_6 = S_5^2$	3
$S_7 = (2, 11, 14)(3, 4, 8)(5, 9, 10)(6, 13, 7)$	2
$S_8 = (1, 13, 12)(9, 4, 14)(3, 8, 6)(5, 10, 7)$	2
$S_9 = (2, 8, 9)(4, 14, 13)(5, 10, 6)(7, 12, 11)$	7
$S_{10} = (2, 7, 8)(3, 10, 11)(5, 13, 9)(6, 12, 14)$	13
$S_{11} = (2, 3, 4)(6, 9, 11)(7, 12, 14)(8, 10, 13)$	7
$S_{12} = S_{11}^2$	7
$S_{13} = (2, 14, 5)(3, 9, 13)(4, 7, 11)(8, 10, 12)$	7
$S_{14} = S_{13}^2$	7
$S_{15} = (1, 10, 6)(3, 8, 9)(4, 11, 12)(7, 13, 14)$	7
$S_{16} = (1, 10, 4)(3, 6, 14)(5, 12, 8)(9, 13, 11)$	13

Since no group satisfying the relations $(2, 3, 6)$ is simple,[10] it is obvious that none of the first eight of the above operators when coupled with T will generate the entire group. S_{12} and S_{14} satisfy with T the same defining relations as do S_{11} and S_{13}, respectively.

S_9 is transformed into S_{15} by the substitution

[10] G. A. Miller, *Quarterly Journal*, vol. 33 (1901-1902), p. 76.

$$(1,2)(3,5)(4,7)(6,9)(8,10)(11,12)(13,14)$$

which is commutative with T. Similarly S_{10} is transformed into S_{16} by

$$(1,7,10,5,9,11,12,6,3,8,4,2)$$

which is also commutative with T. Hence there remain only S_9, S_{10}, S_{11} and S_{13} to be considered. The periods of the commutators of S_9, S_{11} and S_{13} with T are respectively 13, 6, 7. Hence there are four distinct definitions for G_{1092}.

It has already been proved that the relations $(2,3,7;6)$ and $(2,3,7;7)$ are sufficient[11] to determine G_{1092}. We pass then to the definition in terms of $(2,3,7;13)$. That these four relations are not sufficient to define G_{1092} follows from the fact that Professor Brahana has shown that the simple group of order 9828 can also be generated by two operators satisfying the same relations.[12] It becomes necessary then to consider a fifth condition, and it will appear in what follows that the new relation $(Q^2P^6)^3 = 1$ is sufficient to define G_{1092}. Indeed, it will be shown to be sufficiently strong to make the condition $Q^{13} = 1$ redundant.

We first consider some consequences of the initial relations $(2,3,7)$; $(Q^2P^6)^3 = 1$. In order to facilitate the verification of the manipulation which follows, a number has occasionally been written to the right to indicate the particular relation in Art. 2 which is being used in the process of simplification.

$$\begin{aligned} Q^3P^6Q^3P^2Q^{-1}P^2 &= Q^3P^6Q^3P^2 \cdot Q^{-1}P^4 \cdot P^5 \\ &= Q^3P^6Q^2P^6 \cdot PQP^5QP^5. \end{aligned}$$

But
$$\begin{aligned} PQP^5QP^5 &= PQP \cdot P^4QP^4 \cdot P && (1), (4) \\ &= QP^5Q^{-1}P = QP \cdot P^4Q^{-1} \cdot P \\ &= Q \cdot PQP^4Q \cdot Q^{-1} = Q^2P^6Q^{-1} && (3) \\ \therefore\quad Q^3P^6Q^3P^2Q^{-1}P^2 &= Q(Q^2P^6)^3Q^{-1} = 1 \\ Q^3P^6Q^3 &= P^5QP^5. \end{aligned}$$

Now $P^5QP^5 = P^5(QP^3)P^2$ is of period two. Hence Q^6P^6 is also of period two. Therefore

$$\begin{aligned} Q^{-6} &= P^{-1}Q^6P^{-1} = P^{-1}Q \cdot Q^4 \cdot QP^{-1} && (1) \\ &= QPQ^{-1}P^{-2} \cdot Q^4 \cdot P^{-2}Q^{-1}PQ \\ Q^{-8} &= PQ^{-1}P^{-1} \cdot P^{-1}Q^4P^{-2}Q^{-1} \cdot P = P^2 \cdot P^{-1}Q^{-1}P^{-1} \cdot QP^2Q^{-4}P^2 && (2), (1) \\ &= P^2 \cdot Q^{-1}P^{-2}Q^{-1} \cdot QP^2Q^{-4}P^2 = P^2Q^{-5}P^2 \\ 1 &= QP^2Q^8P^2Q^{-6} = QP^2Q \cdot Q^7P^2Q^{-6} && (1) \\ &= P \cdot QPQ^7P^2 \cdot Q^{-6} = P^{-1}Q^{-7}P^{-1}Q^{-7} && (2) \\ (Q^7P)^2 &= 1. \end{aligned}$$

---

[11] "Certain perfect groups," p. 354; A. Sinkov, *Bulletin of the American Mathematical Society*, vol. 41 (1935), p. 239.

[12] "Pairs of generators," p. 532.

It now follows from the results at the end of Art. 2 that $Q^{13} = 1$ is a consequence of the initial relations. It then results from Dr. Coxeter's work that $(Q^8P^2)^2 = 1$ and that the group in question is either $G^{3,7,13}$ or a quotient group of it.

With the aid of these further relations, it becomes a simple matter to prove that the conditions $(2, 3, 7)$; $(Q^2P^6)^3 = 1$ define G_{1092}. The procedure and the notation are the same as those used by Professor Brahana[13] in his study of $(2, 3, 7; 6)$.

Every combination of P and Q is reducible to an operator in some one of the 156 co-sets obtainable from those listed below.

$$\begin{array}{lll} a = 1 & e = Q^2P & i = Q^2P^6 \\ b = QP & f = Q^2P^2 & j = Q^3P \\ c = QP^4 & g = Q^2P^4 & k = Q^3P^2 \\ d = QP^5 & h = Q^2P^5 & l = Q^3P^4. \end{array}$$

The representation of P and Q as permutations of these 156 symbols is as follows:

$$Q = (aa_1a_2 \cdots a_{12})(bb_1b_2 \cdots b_{12}) \cdots (ll_1l_2 \cdots l_{12})$$
$$P = (abb_{12}a_{12}cdc_1)(a_2efc_{12}ghi)(a_3jkg_{12}lh_4d_{10})$$
$$(a_4j_9e_2l_{12}k_9d_8i_9)(a_5e_5k_4k_8e_7a_8c_7)(a_6b_5f_7e_6f_8b_7a_7)$$
$$(a_9i_6d_5k_3l_5e_{10}j_7)(a_{10}d_3h_5l_4g_2k_{12}j_3)(a_{11}i_2h_9g_1c_2f_2e_{12})$$
$$(b_1f_1b_{11}d_{12}i_{12}i_3d_1)(b_2k_{11}e_{11}i_{11}g_{10}d_4h_{10})(b_3e_9j_2g_9l_{10}i_7h_6)$$
$$(b_4k_7j_6l_9f_9c_8d_6)(b_8d_7c_6f_6l_8j_{10}k_5)(b_9h_3i_8l_7g_5j_1e_3)$$
$$(b_{10}h_{12}d_9g_4i_4e_1k_1)(c_3j_4l_1i_1l_3j_{12}c_{11})(c_4h_1f_3l_2f_{12}h_8c_{10})$$
$$(c_5j_5f_4f_{11}j_{11}c_9g_7)(d_2k_2k_{10}d_{11}g_3j_8g_{11})(e_4l_{11}h_{11}l_6e_8i_{10}i_5)$$
$$(f_5f_{10}h_7g_6k_6g_8h_2)(a)(b_6).$$

As a corollary to this definition of G_{1092} we will now determine the order of $G^{3,7,13} \equiv (2, 3, 7)$; $(Q^7P)^2 = 1$. We shall prove first that these conditions imply $(Q^2P^6)^6 = 1$.

$$\begin{aligned} PQ^2PQ^{-2} &= PQ \cdot QP^3Q \cdot Q^{-1}P^5Q^{-2} \\ &= PQP^4Q \cdot Q^{-2}P^5Q^{-2} = QPQ^2P^2 && (3),\ (5) \\ &= Q^2P^6 \cdot PQ^{-1}P \cdot Q^2P^2 = Q^2P^6Q^{-1} \cdot P^3Q \cdot P^2 && (3) \\ &= Q^2P^6Q^{-2}P^6 \end{aligned}$$

whence $(PQ^2PQ^{-2})^2 = 1$. From this result, we have

$$\begin{aligned} P^6Q^2P^6 &= Q^2PQ^{11}PQ^2 \\ Q^2P^6Q^2P^6Q^2 &= Q^4PQ^{11}PQ^{11} \cdot Q^6 = Q^4PQ^{11}PQ^{11}PQ^7P \\ (Q^2P^6)^3 &= Q^4(Q^2P^6)^{-3}Q^9 \end{aligned}$$

---

[13] "Certain perfect groups," pp. 352, 353.

that is, $(Q^2P^6)^3$ is transformed into its inverse by Q^4. It is therefore commutative with Q and, as a result $(Q^2P^6)^3 = (Q^2P^6)^{-3}$; $(Q^2P^6)^6 = 1$.

We see then that G_{1092} is a factor group of $G^{3,7,13}$, obtained by adjoining the additional restriction [14] $(Q^2P^6)^3 = 1$. The invariant subgroup which gives rise to G_{1092} as a quotient group is the group generated by the complete set of conjugates involving $(Q^2P^6)^3$.

Now $(Q^2P^6)^3$ is invariant under transformation by Q and is therefore equal to $(P^6Q^2)^3$. It is also invariant under transformation by P. For, $P(P^6Q^2)^3P^{-1} = (Q^2P^6)^3$. Therefore $(Q^2P^6)^3$ is invariant in $G^{3,7,13}$, and the order of this perfect group is consequently 2184.

A complete definition for G_{1092} in terms of $(2,3,13;\,7)$ can be readily obtained from the definition based on $(2,3,7;\,13)$. We have already seen that G_{1092} may be defined by

$$C:\quad P^7 = Q^{13} = (QP^2)^3 = (QP^3)^2 = (Q^7P)^2 = (Q^2P^6)^3 = 1.$$

Let us now introduce the substitution $P^2 = \bar{Q}$; $Q = \bar{P}^2$ and let us require that $\bar{P}^{13} = 1$. Then $P = \bar{Q}^4$; $\bar{P} = Q^7$. The defining relations become

$$D:\quad \bar{P}^{13} = \bar{Q}^7 = (\bar{Q}\bar{P}^2)^3 = (\bar{Q}^5\bar{P}^2)^2 = (\bar{Q}^4\bar{P})^2 = (\bar{Q}^3\bar{P}^4)^3 = 1,$$

in which the roles played by P and Q in C have been interchanged. Since the additional condition $(Q^2P^6)^3 = 1$ has been replaced by $(Q^3P^4)^3 = 1$ it follows that G_{1092} is completely defined by $(2,3,13)$; $(Q^3P^4)^3 = 1$. We have then the following necessary and sufficient condition:

THEOREM. *Two operators of periods* 2 *and* 3 *generate the simple group of order* 1092 *if and only if they satisfy one of the following sets of independent relations*

$$\begin{aligned} A:&\quad (2,3,7;\,6)\\ B:&\quad (2,3,7;\,7)\\ C:&\quad (2,3,7);\quad (Q^2P^6)^3 = 1\\ D:&\quad (2,3,13);\quad (Q^3P^4)^3 = 1. \end{aligned}$$

5. *The simple group of order* 5616. It is easily verified by direct calculation with the generating permutations given by Bisshopp that the period of the commutator of S and T is four for one pair of generators (the pair for which $\bar{S} = R^{-1}SR$) and six for the other $(\bar{S} = R^{-9}SR^9)$. Hence G_{5616} is a factor group of $(2,3,13;\,4)$ and $(2,3,13;\,6)$. We consider first the case $Q^4 = 1$ and proceed to show that one further restriction on P and Q is sufficient

[14] Dr. Coxeter has found that this additional restriction is equivalent to $(CBA)^7 = 1$, where the operators A, B, and C are the generators of $G^{3,7,13}$ in his definition.

to define G_{5616}. This additional relation is $QP^7QP^7 = P^6Q^2P^7Q^2$ which implies $(QP^9)^4 = 1$. For,

$$QP^7QP^7 = QP^7 \cdot P^{10}Q^3P^{10} \cdot P^7 = QP^4Q^3P^4$$
$$Q^3P^4Q^3P^4 = Q^2P^6Q^2P^7Q^2 = (P^7Q^2)^{-1}Q^2(P^7Q^2).$$

This in turn implies $(QP^{10})^4 = 1$. For,

$$QP^9 = QP^3 \cdot P^6 = P^{10}Q^3P^6 = P^7(P^3Q^3)P^6.$$

The period of $Q^{-1}P^7$ is also four since $Q^{-1}P^7 = Q^{-1}P^{-3} \cdot P^{10} = P^3QP^{10}$. The proof of sufficiency will consist in showing that the five relations chosen can be used to establish all of the conditions given by Bisshopp. The reduction of these conditions to expressions involving P and Q will be omitted as that involves no particular difficulty. The following reductions will however be found useful in simplifying the process.

$$\begin{aligned} \bar{S} &= P^2Q^2 \\ \bar{S}T &= P^2QP^{10} \\ (\bar{S}T)^4 &= P^3Q^2P^9QP^{12}Q \\ \bar{S}^2T\bar{S}T &= QP^{11}Q^3P^9 \end{aligned}$$

(1) $(P^2QP^{10})^8 = 1$

$$\begin{aligned} (P^2QP^{10})^2 &= P^2QP^2 \cdot P^{10}QP^{10} \\ &= Q^3P^{11} \cdot Q^3P^{10} \cdot QP^{10} \\ &= Q^3PQ^2 \cdot P^{10}Q^3 \cdot Q \\ &= Q^3PQ^3P^3Q \\ &= (P^3Q)^{-1}P^4Q^3(P^3Q) \end{aligned}$$

(2)
$$\begin{aligned} &QP^{11}Q^2P^4Q^3P^2Q^3P^{11}Q^2 = P^3Q^3P^2QP^4Q^3P^6Q^2P^2 \\ &QP^{11}Q^2P^4Q^3P^2Q^3 \cdot P^{11}Q^2P^{11}Q^2 \cdot P^7QP^9Q^3P^{11}QP^{10} = 1 \qquad (5) \\ &QP^{11}Q^2P^4Q^3P^2 \cdot QP^9QP^9 \cdot Q^3P^{11}QP^{10} = 1 \qquad (5) \\ &QP^{11}Q^2P^4 \cdot Q^3P^6Q^3P^6 \cdot Q^2P^5Q = 1 \\ &Q \cdot P^{11}Q^2P^{11}Q^2P^{11} \cdot Q = Q^4 = 1 \end{aligned}$$

(3) $(QP^{11}Q^3P^9)^6 = 1$

$$\begin{aligned} (QP^{11}Q^3P^9)^3 &= QP \cdot QP^9 \cdot Q^3P^{11}Q^2P^5Q \\ &= QP^5Q^3P^4Q^3P^4 \cdot Q^2P^{11}Q^2 \cdot P^5Q \qquad (5) \\ &= QP^5Q^3P^4Q^3 \cdot P^6Q^2P^7Q^2 \cdot Q^3 \\ &= QP^5Q^3P^{12} \cdot QP^3 \cdot PQP^8Q^3 \\ &= (PQP^8Q^3)^{-1}QP^3(PQP^8Q^3) \end{aligned}$$

(4′)
$$P^5Q^2P^2QP^9 = P^4QP^2Q^2P^{10}$$
$$P^5Q^2P^2 \cdot QP^{11} \cdot Q^2P^{11}Q^3P^9 = 1 \qquad (3)$$
$$P^5Q^2 \cdot P^3Q \cdot P^4 \cdot Q^3P^{11}Q^3 \cdot P^9 = 1$$
$$P^5QP^3QP^{11} = P^{26} = 1$$

(5′)
$$P^5Q^2P^5Q^2P^{10}QP^5Q = PQ^3PQ^2P^{10}Q$$
$$QP^4Q^2P^5Q^2P^{10}QP^8Q^2P^{12} = 1$$
$$QP^2 \cdot P^2Q^2P^2 \cdot P^3Q \cdot QP^{10}QP^8Q^2P^{12} = 1 \qquad (5)$$
$$P^{11}Q^3P^{11}QP^{11} \cdot QP^{10}QP^{10}QP^{10} \cdot P^{11}Q^2P^{12} = 1$$
$$P^{11}Q^3P^{11}QP \cdot Q^3P^{11} \cdot Q^2P^{12} = 1$$
$$P^{11}Q^3P^{11} \cdot QP^3Q \cdot P^2Q^3P^{12} = 1$$
$$P^{11}Q^3P^{10}Q^3 \cdot P^{12} = 1$$
$$P^{26} = 1$$

(6′)[15]
$$P^5Q^3P^9Q^2P^{10}Q^2P^{11} = Q^3P^7QP^9Q$$
$$P^5Q^3P^9Q^2P^{10}Q^2 \cdot P^{11}Q^3 \cdot P^4Q^3P^6Q = 1$$
$$P^5Q^3P^9Q^2 \cdot P^{10}Q^3 \cdot P^2QP^6Q^3P^6Q = 1$$
$$P^9 \cdot P^9Q^3P^9Q^3 \cdot P^5QP^6Q^3P^6Q = 1$$
$$P^9QP^9Q \cdot P^6Q^3P^6Q = 1$$
$$Q^3P^4 \cdot Q^3P^{10}Q^3 \cdot P^6Q = 1$$
$$Q^4 = 1.$$

It follows then that G_{5616} is completely defined by the relations

$$(2, 3, 13;\ 4);\ QP^7QP^7 = P^6Q^2P^7Q^2.$$

In the same manner as above, it can be demonstrated that one additional restriction $Q^2P^6Q^2 = P^2Q^5P^2$ adjoined to $(2, 3, 13;\ 6)$ is sufficient to define G_{5616}. The manipulation is considerably more involved but the procedure is essentially the same.

We have then the following theorem:

Two operators of periods two and three generate G_{5616} if and only if they satisfy one of the following sets of relations

$$A:\quad (2, 3, 13; 4); \qquad QP^7QP^7 = P^6Q^2P^7Q^2$$
$$B:\quad (2, 3, 13; 6); \qquad Q^2P^6Q^2 = P^2Q^5P^2.$$

The author wishes to express here his extreme indebtedness to Dr. Coxeter for the many valuable suggestions and criticisms received in the course of this research.

WASHINGTON, D. C.

---

[15] N. B. There is a typographical error in the statement of this relation. The right member should be $\bar{S}(\bar{S}T)^4\bar{S}(\bar{S}T)^3\bar{S}^2$.

HYPERGROUPS.[1]

By H. S. Wall.

A hypergroup is a system in which any two elements a, b can be combined to form the product ab, which is a complex of n not necessarily distinct elements of the system. Here n is a fixed integer $\geqq 1$. If $n = 1$ the hypergroup reduces to an ordinary group. If $[ab]$, called the bracket product, is the set of all the distinct elements of ab, the totality of elements a such that $[ax]$ and $[xa]$ are single elements for every x forms a group with respect to the bracket product. This group is called the nucleus. A hypergroup in which the nucleus is not vacuous is related to the Mischgruppe of Loewy and Baer (cf. § 3), and is called a hypergroup of type M.

This paper is in three parts. Part I contains a postulational basis for abstract hypergroups. In Part II there is a discussion of isomorphism and automorphism for hypergroups, and the notion of conjugate sets in groups is extended to hypergroups of type M. Part III contains a study of special hypergroups, principally multiplicative systems derived from groups.

PART I.

Abstract Hypergroups.

1. *Postulational basis and definitions.* A *hypergroup* is a system H of distinct *elements* in which there is defined an operation called *multiplication.* With each hypergroup there is associated an integer $n \geqq 1$ called the *dimension* of the hypergroup.

I. *The product postulate.* If a, b are two elements of H, not necessarily distinct, then the *product ab* is a complex of n uniquely determined elements of H. We shall write for convenience

$$ab = c_1 + c_2 + \cdots + c_n = \sum_{i=1}^{n} c_i.$$

The order in which the elements are written in the "sum" is immaterial, but ab is in general different from ba. The n elements of the product ab are not necessarily distinct.

[1] Received December 28, 1935; revised July 24, 1936.

Let $A = a_1 + a_2 + \cdots + a_p$, $B = b_1 + b_2 + \cdots + b_q$ be two complexes of elements of H. Then we shall define the product AB to be the complex

$$\sum_{i=1}^{p} \sum_{j=1}^{q} a_i b_j.$$

II. *The associative law.* If a, b, c are three elements of H, then

$$a(bc) = (ab)c = abc.$$

The members of the last equality have a meaning by the preceding definition of the product of two complexes.

III. *The identity postulate.* There is at least one element e in H, called an *identity* element, such that for every element a in H the products ae and ea both contain the element a at least once.

IV. *The postulate of the inverse.* There is an identity element e in H such that corresponding to each element a in H there is at least one element a^{-1} in H such that the products aa^{-1} and $a^{-1}a$ both contain e at least once. The element a^{-1} is called an *inverse* of a.

Evidently a hypergroup of dimension unity is a group.

The *bracket product* $[ab \cdots c]$ of two or more elements of H will be defined to be the complex of all the distinct elements in the product $ab \cdots c$. An element a of H such that $[ax]$ and $[xa]$ are single elements for every x in H will be called a *scalar*. The set of all the scalars of H will be called the *nucleus*. A hypergroup in which the nucleus contains at least one element is a *hypergroup of type M*. It will appear later that a hypergroup of type M is closely related to the *Mischgruppe* of Loewy and Baer.

We shall employ the notation nG ("n-group") for a hypergroup of dimension n; and nGM for an nG of type M. An nG_1 is an nG containing just one identity element, and in which each element a has just one inverse a^{-1} such that if aa' or $a'a$ contains e then $a' = a^{-1}$. An nG_1M is an nG_1 of type M. The notation $a \varepsilon K$, where a is an element and K is a set of elements, signifies that a is contained in K.

Since $b \varepsilon a(a^{-1}b)$ it follows that there is an element $x \varepsilon a^{-1}b$ such that $b \varepsilon ax$. Likewise, there is an element $y \varepsilon ba^{-1}$ such that $b \varepsilon ya$. We therefore have this result.

THEOREM 1. *If a, b are any two elements of an nG, H, then H contains two elements x and y such that*

(1) $$b \varepsilon ax \quad \textit{and} \quad b \varepsilon ya.$$

Suppose that $x_1, x_2, \cdots, x_{n+1}$ are $n+1$ different values of x such that $b \,\varepsilon\, ax$. Then

$$e \,\varepsilon\, b^{-1}ax_i, \qquad (i = 1, 2, \cdots, n+1).$$

Thus there is at least one of the n elements in the product $b^{-1}a$ which has more than one inverse.

THEOREM 2. *If in Theorem 1 H is an nG_1, then there are at most n elements x and at most n elements y satisfying* (1).

The upper bound given by this theorem for the number of solutions of (1) is actually attainable as will be seen from examples to be given later.

THEOREM 3. *Let H be an nG of finite order in which the relation $b \,\varepsilon\, ax$ has just one solution x for each pair of elements a, b. Then all the elements of H are scalars.*

Proof. If some product ac contains two or more distinct elements then, under the hypothesis, there must be an element d for which the product ad contains no elements. But this is contrary to postulate I.

From this theorem and Theorem 4 below it follows that an nG of finite order in which the relation $b \,\varepsilon\, ax$ has a unique solution x for every a and b is an ordinary group with respect to the bracket product.

THEOREM 4. *The nucleus of an nGM is a group with respect to the bracket product.*

Proof. Let H contain a scalar element δ. Then

$$[\delta\delta^{-1}] = [\delta^{-1}\delta] = e$$

where e is an identity element. Let x be any element of H, and determine y in accordance with Theorem 1 so that $x \,\varepsilon\, \delta y$, or, since δ is a scalar, $x = [\delta y]$ and $[\delta^{-1}x] = [ey]$. Now $[\delta e y] = [\delta y] = a$ (say) and hence $[\delta\delta^{-1}x] = [ex] = a$, so that $[ex]$ is a single element of H for every x in H. In like manner we find that $[xe]$ is a single element of H for every x in H. Suppose now that H contains another identity e'. Then $[ee'] = e = e'$. Thus *an nGM contains a unique identity element e, and e is a scalar.*

It follows from the preceding discussion that if δ is a scalar, then δ^{-1} is a scalar; and it is easy to see that δ can have but one inverse.

Let σ, δ be two scalars, and put $[\sigma\delta] = c$. Then

$$[\sigma^{-1}\sigma\delta] = [\sigma^{-1}c], \quad [e\delta] = [\sigma^{-1}c], \quad [\sigma^{-1}c] = \delta.$$

Choose y in accordance with Theorem 1 so that $x \varepsilon y\sigma^{-1}$, where x is any assigned element of H. Then

$$[y\delta] = [y\sigma^{-1}c] = [xc]$$

is a single element of H. In like manner we find that $[cx]$ is a single element of H. Thus c is a scalar. This completes the proof of the theorem.

An nGM is not necessarily an nG_1. For consider the following example.

	e	a	b
e	$e+e$	$a+a$	$b+b$
a	$a+a$	$e+b$	$e+b$
b	$b+b$	$e+a$	$e+a$

This is a $2GM$ but not a $2G_1$. It is noteworthy that $aa = ab$, $a \neq b$.

2. *Properties of hypergroups of type M.* Let H be an nGM and suppose that

$$(2) \qquad ab_1 = ab_2 = \cdots = ab_{n+1}, \qquad b_i \neq b_j,\ i \neq j.$$

Then $a^{-1}ab_1 = a^{-1}ab_2 = \cdots = a^{-1}ab_{n+1}$. Now $a^{-1}ab_{n+1}$ contains n^2 elements of which n are the element b_{n+1} repeated n times. Among the remaining $n(n-1)$ elements there must be the elements $b_1, b_2, \cdots, b_n$ each repeated n times. But this is impossible. Hence we have this result.

THEOREM 5. *If H is an nGM then the relation* (2) *is impossible.*

Of course a relation obtained by reversing the orders of the factors in (2) is impossible in an nGM.

THEOREM 6. *The only element a of an nGM such that $[aa] = a$ is the identity element e.*

Proof. Put $a^{-1}a = e + c_2 + c_3 + \cdots + c_n$. Then since

$$aa = a + a + \cdots + a = na, \qquad a^{-1}aa = na^{-1}a,$$

we have

$$na + (c_2 + c_3 + \cdots + c_n)a = ne + n(c_2 + c_3 + \cdots + c_n).$$

Now if $a \neq e$, then $a = c_2$ (say) and hence

$$na + (c_3 + c_4 + \cdots + c_n)a = ne + n(c_3 + c_4 + \cdots + c_n).$$

By repeated application of this argument we shall arrive finally at the equality

$$na + c_n a = ne + nc_n,$$

which is impossible if $a \neq e$.

THEOREM 7. *If in an nGM* $[ab]$ *and* $[bc]$ *are single scalars, then* b *is a scalar.*

Proof. Determine y and z so that $x \varepsilon ya$ and $x \varepsilon cz$. Then $[xb]$ and $[bx]$ are single elements. Since x is arbitrary, b is a scalar.

COROLLARY. *If in an nGM* $[b'b] = [bb''] = e$, *the identity, then* b *is a scalar and* $b' = b'' = b^{-1}$.

THEOREM 8. *Let* a *be an element of an* nG_1M *such that* $[aa^{-1}] = e$. *Then* $[a^{-1}a] = e$.

Proof. Let x be any element of an nG_1M, H. Then by Theorem 1 there is an element y in H such that $x \varepsilon a^{-1}y$. Now $[aa^{-1}] = e$, $[aa^{-1}y] = y$, and therefore

$$[ax] = y. \tag{3}$$

Likewise we find that

$$[xa^{-1}] = z, \tag{4}$$

a single element of H. As special instances of (3) we have

$$[aa] = b, \qquad [ab^{-1}] = c, \qquad [ac] = d, \tag{5}$$

all single elements of H. Hence $[aab^{-1}] = [bb^{-1}] = [ac] = d$. Thus

$$[bb^{-1}] = e, \tag{6}$$

so that b is an element which satisfies the hypothesis of our theorem.

Put

$$b^{-1}a = a^{-1} + t_2 + t_3 + \cdots + t_n,$$
$$b^{-1}b = e + s_2 + s_3 + \cdots + s_n,$$
$$a^{-1}a = e + \sigma_2 + \sigma_3 + \cdots + \sigma_n.$$

Then from the first equality (5) we obtain $b^{-1}aa = nb^{-1}b$, or

$$e + \sigma_2 + \sigma_3 + \cdots + \sigma_n + (t_2 + t_3 + \cdots + t_n)a = ne + n(s_2 + s_3 + \cdots + s_n). \tag{7}$$

This holds if $n > 1$. Of course our theorem is granted if $n = 1$.

Assume now that the theorem is false. Then there must be an integer k, $0 < k < n$, such that for every element w satisfying the equation $ww^{-1} = ne$;

$w^{-1}w$ contains the element e at least k times; and such that there is at least one value a of w for which $a^{-1}a$ contains e just k times. Then if the notation is properly chosen, (7) takes the form

$$ke + \sigma_{k+1} + \sigma_{k+2} + \cdots + \sigma_n + (t_2 + t_3 + \cdots + t_n)a$$
$$= nke + n(s_{k+1} + s_{k+2} + \cdots + s_n),$$

where $\sigma_i \neq e$, $i = k+1, k+2, \cdots, n$. But this implies that $t_i = a^{-1}$, $i = 2, 3, \cdots, n$, or, $[b^{-1}a] = a^{-1}$. Determine x so that $a^{-1} \varepsilon\, xb^{-1}$, apply (4), and we get

$$[xb^{-1}a] = [xa^{-1}] = z,$$

and consequently $a^{-1}a = ne$, contrary to the assumption concerning a.

Corollary. *If a is an element of an nG_1M such that $[aa^{-1}] = e$, then a is a scalar.*

This is a consequence of Theorems 7 and 8.

Theorem 9. *If a, b are elements of an nG_1M such that $[ab]$ is a single scalar, then a and b are scalars.*

Proof. Determine x so that $a^{-1} \varepsilon\, bx$; then $[ab] = c$, $[abx] = [cx] = d$, $[aa^{-1}] = d = e$. Hence by the above corollary, a is a scalar. In like manner we find that b is a scalar.

An important consequence of this discussion concerns subhypergroups. A subset K of elements of H is a *subhypergroup* of H if the elements of K satisfy the postulates for a hypergroup. We have this theorem.

Theorem 10. *If the identity e of an nGM H is in K, a subhypergroup of H, then K is an nGM whose nucleus is a subgroup of the nucleus of H.*

Proof. Since K contains the scalar e, K is an nGM and e is the identity of K. Let a be any scalar of K and a^{-1} an inverse of a in K. Then $[aa^{-1}] = [a^{-1}a] = e$, so that by Theorem 7 a is a scalar of H.

3. *The Mischgruppe of Loewy*[2] *and Baer.*[3] A *Mischgruppe* M is a system in which the product km is defined when and only when k is in a subset K of M called the *Kern*. The Mischgruppe may be defined by the following four postulates.

I. The Kern K is an ordinary group.

[2] Loewy, *Crelle*, Bd. 157 (1927), pp. 239-254.

[3] Baer, *Sitzungsbericht Heidelberger Akad. Wiss. Math-nat. Kl.* (1928) (4).

II. If a and b are any two elements of M then the product ab is defined if and only if a is in K; and when defined, ab is a uniquely determined element of M.

III. If a and b are in K, and c is in M, then

$$(ab)c = a(bc).$$

IV. If a is in K and b is in M then $ab = b$ if and only if a is the identity element of the group K.

Baer omitted the explicit statement of IV although it is clearly independent of I, II, III as the following example shows.

	e	a	A	B	C	D	E
e	e	a	A	B	C	D	E
a	a	e	D	E	C	A	B

Given a Michgruppe M which is not of order 2 nor a group. Then there is determined (not in general uniquely) a group G with subgroup E such that M is simply isomorphic with the set of all distinct cosets

$$M': \qquad Eg_1, Eg_2, Eg_3, Eg_4, \cdots; \; Eh_1, Eh_2, Eh_3, \cdots$$

where $g_1, g_2, g_3, \cdots$ are commutative with E, and $h_1, h_2, h_3, \cdots$ are not commutative with E. In M' multiplication is defined by the formula

$$(Eg)(Ea) = (Ega).$$

This holds if and only if g is commutative with E: $Eg = gE$. If we put

$$(Ea)(Eb) = \sum_{x \text{ in } E} (Eaxb)$$

then with this definition of multiplication M' is a hypergroup of type M whose nucleus (with respect to the bracket product) is the Kern of the Mischgruppe. Conversely, in an nGM let us identify the nucleus with the kern K; and define the symbol "o" as follows. Let $a \circ b = [ab]$ if and only if a is in the nucleus. Then the system so defined satisfies postulates I, II, III of the Mischgruppe. Postulate IV is not in general satisfied.

We have seen that multiplication can be defined between arbitrary elements of a Mischgruppe in such a way that the new system so obtained in a hypergroup of type M. This can be done in more than one way. For consider the Mischgruppe given by the following table:

	e	a	A	B
e	e	a	A	B
a	a	e	B	A.

To this there correspond the following two hypergroups of dimension 2.

	e	a	A	B
e	$e+e$	$a+a$	$A+A$	$B+B$
a	$a+a$	$e+e$	$B+B$	$A+A$
A	$A+A$	$B+B$	$e+B$	$a+A$
B	$B+B$	$A+A$	$a+A$	$e+B$

	e	a	A	B
e	$e+e$	$a+a$	$A+A$	$B+B$
a	$a+a$	$e+e$	$B+B$	$A+A$
A	$A+B$	$A+B$	$e+a$	$e+a$
B	$B+A$	$B+A$	$a+e$	$e+a$.

The first is a $2G_1M$ while the second is not of this character.

4. *The identity groups.* Let E_a denote the set of elements of the nucleus of an nGM, H, such that $[E_a a] = a$. It is clear that E_a contains at least one element, namely the identity element; and that E_a is a group with respect to the bracket product. We shall call E_a the left a-identity group. Similarly, E'_a is the right a-identity group. It is the subgroup of the nucleus of H such that $[aE'_a] = a$. E_a and E'_a may be of different orders.

THEOREM 11. *If H is an nGM, and a any element of H, then E_a is contained in aa^{-1} and E'_a is contained in $a^{-1}a$.*

Proof. Let δ be an element of E_a. Then $[\delta a] = a$, $[\delta aa^{-1}] = [aa^{-1}]$. Hence since $\delta \varepsilon \delta aa^{-1}$ it follows that $\delta \varepsilon aa^{-1}$. Likewise we find that if δ is in E'_a then $\delta \varepsilon a^{-1}a$.

THEOREM 12. *If H is an nG_1M, and a any element of H, and we put*

$$[aa^{-1}] = E_a + F_a,$$
$$[a^{-1}a] = E'_a + F'_a,$$

then F_a and F'_a contain no scalars.

Proof. It will suffice to prove that if δ is any scalar $\varepsilon\, aa^{-1}$ then δ is in E_a. Write

$$aa^{-1} = e + \delta + S.$$

Then $\delta^{-1}aa^{-1} = n\delta^{-1} + ne + \delta^{-1}S$, and consequently $[\delta^{-1}a] = a$ so that δ^{-1} is in E_a. Therefore, since E_a is a group, $\delta \varepsilon E_a$, as was to be proved.

THEOREM 13. *In an nG_1M let an element δ of E_a be repeated just r times in aa^{-1}. Then every element of E_a is repeated just r times in aa^{-1}.*

Proof. By hypothesis $aa^{-1} = r\delta + S$, where S does not contain δ. Then

$$aa^{-1}\delta^{-1} = nre + T$$

where T does not contain e. Hence $a^{-1}\delta^{-1} = na^{-1}$, $naa^{-1} = nre + T$, so that aa^{-1} contains e just r times. Now if aa^{-1} contains an element δ' of E_a just s times, $s \neq r$, then the preceding argument would show that aa^{-1} contains e just s times, which is impossible.

COROLLARY. *If in Theorem 13 E_a contains an element repeated just r times in aa^{-1}, then for every x there is a y such that ay contains x repeated at least r times.*

Proof. By Theorem 13, aa^{-1} contains e repeated just r times. Hence $aa^{-1}x$ contains x at least nr times. But this is impossible unless $a^{-1}x$ contains an element y such that ay contains x at least r times.

THEOREM 14. *If a and b are elements of an nG_1M such that*

$$[ab] = \delta_1 + \delta_2 + \cdots + \delta_k + S \qquad (k \geqq 1)$$

where $\delta_1, \delta_2, \cdots, \delta_k$ are scalars and S contains no scalars, then

$$E_a = [\delta_i\delta_1^{-1}] + [\delta_i\delta_2^{-1}] + \cdots + [\delta_i\delta_k^{-1}], \qquad (i = 1, 2, \cdots, k).$$

Proof. It follows from the hypothesis that

$$\delta_1^{-1}a = \delta_2^{-1}a = \cdots = \delta_k^{-1}a \ (= nb^{-1}).$$

Consequently $[\delta_i\delta_j^{-1}]$ is in E_a. Conversely, let δ be in E_a. Then $[\delta a] = a$, $[\delta ab] = [ab] = \delta_1 + \delta_2 + \cdots + \delta_k + S$, $[\delta_i^{-1}\delta a] = b^{-1}$. Put $[\delta_i^{-1}\delta] = \sigma$. Then $[\sigma ab] = e + T$ (say), $[ab] = \sigma^{-1} + U$, and therefore there is a j such that $\sigma^{-1} = \delta_j$, $[\delta^{-1}\delta_i] = \delta_j$, $\delta = [\delta_i\delta_j^{-1}]$, which completes the proof of the theorem.

COROLLARY. *If in an nG_1M, E_a contains just k elements, then if $[ab]$ contains one scalar it must contain just k scalars.*

Let B_a denote the set of elements of the nucleus of an nGM which are commutative with a. Clearly B_a is a group with respect to the bracket product; and any element which is in two of the groups B_a, E_a, E'_a is necessarily in the third also.

If b is in B_a, then

$$[(b^{-1}E_ab)a] = [b^{-1}E_aab] = [b^{-1}ab] = a,$$
$$[a(b^{-1}E'_ab)] = [b^{-1}aE'_ab] = [b^{-1}ab] = a.$$

That is, $[b^{-1}E_ab] = E_a$, $[b^{-1}E'_ab] = E'_a$. In particular we have this result.

THEOREM 15. *In an nGM the elements common to E_a and E'_a form an invariant subgroup of B_a.*

Let $\Delta = E_a + [\delta_2 E_a] + [\delta_3 E_a] + \cdots$ be a decomposition of the nucleus Δ of an nGM, H, into distinct cosets. Then

$$[\Delta a] = a + [\delta_2 a] + [\delta_3 a] + \cdots$$

is a decomposition of the complex $[\Delta a]$ into distinct elements. If H is an nG_1M and we put $e = \delta_1$, $a_i = [\delta_i a]$, then it is easy to show that

$$[a_i^{-1}a_i] = [a_j^{-1}a_j], \qquad (i, j = 1, 2, 3, \cdots);$$
$$[\delta_i^{-1}a_ia_i^{-1}\delta_i] = [aa^{-1}], \qquad (i = 1, 2, 3, \cdots).$$

5. *Matrix representation of hypergroups.* Let $H(a_1, a_2, \cdots, a_k)$ be an nG of order k. Then there exists a set of k by k matrices in one to one correspondence with the elements of H in such a way that when

$$a_ia_j = a_{s_1} + a_{s_2} + \cdots + a_{s_n},$$

then

$$A_iA_j = A_{s_1} + A_{s_2} + \cdots A_{s_n},$$

where $A_\nu \sim a_\nu$, $\nu = 1, 2, 3, \cdots, k$.

The matrices A_ν may be formed as follows. Let $x_1, x_2, \cdots, x_k$ be positive integers, and put $X = \sum_{i=1}^{k} x_ia_i$, $X_j = a_jX = \sum_{i=1}^{k} x_i{}^ja_i$, where the $x_i{}^j$ are positive integers of the form

$$x_i{}^j = \sum_{r=1}^{k} p^j{}_{ri}x_r, \qquad (i, j = 1, 2, 3, \cdots, k),$$

in which $p^j{}_{ri} \geq 0$ is an integer. The matrices $A_j = (p^j{}_{ri})$, $j = 1, 2, 3, \cdots, k$, will be seen to have the required property. For, on the one hand,

$$a_ia_jX = (a_{s_1} + a_{s_2} + \cdots + a_{s_n})X = \sum_{r=1}^{k} X_{s_r} = \sum_{r=1}^{k} (x_r{}^{s_1} + x_r{}^{s_2} + \cdots + s_r{}^{s_n})a_r$$
$$= \sum_{r=1}^{k} x'_ra_r,$$

where $x'_r = \sum_{t=1}^{k} (p^{s_1}_{tr} + p^{s_2}_{tr} + \cdots + p^{s_n}_{tr})x_t$; while on the other hand,

$$a_ia_jX = a_iX_j = \sum_{r=1}^{k} x_r''a_r \quad \text{where} \quad x_r'' = x_r' = \sum_{t=1}^{k} (\sum_{u=1}^{k} p^i{}_{tu}p^j{}_{ur})x_t.$$

Thus we have the matrix equation $A_iA_j = A_{s_1} + A_{s_2} + \cdots + A_{s_n}$, as was to be proved.

The matrix representation is related to the following more general kind of representation suggested to me by Dr. Baer. Let H be a *finite* or *infinite* nG. Denote by H^* the set of all finite subsets of H. If a is an element of H, S an element of H^*, then denote by $f_a(S)$ the set of all the elements x of H which are contained in sets sa with s in S. Since n is finite $f_a(S)$ is finite, and $f_a(S)$ is therefore a single-valued function defined for all elements of H^* and mapping H^* into itself. If a and b are two elements of H, then

$$f_b(f_a(S)) = \sum_{x \text{ in } ab} f_x(S).$$

This representation of the hypergroup is analogous to the well-known Cayley representation of groups as permutation groups.

PART II.

Conjugate Sets and Isomorphism.

6. *Complete sets of conjugates in an nGM.* Let H be an nGM with nucleus Δ. Let $a_1 \varepsilon H$, and let

$$[a_1] \tag{8}$$

be the different elements of H of the form $[\delta^{-1}a_1\delta]$ where δ is in Δ. Any element of (8) is called a *conjugate* of any other element of (8). The set is the *complete set of conjugates* of any one of its members. If b_1 is any element of H not in (8) then the complete set of conjugates of b_1 has no elements in common with (8).

Theorem 16. *In an nGM let the nucleus Δ be of finite order q, and B_a, the subgroup of all elements of Δ commutative with a, of order p. Then the number of elements in the complete set of conjugates of a is q/p.*

Proof. Let $\delta_1, \delta_2, \cdots, \delta_p$ be the elements of B_a, and let δ be any element of Δ not in B_a. Then

$$[\delta_1\delta], [\delta_2\delta], \cdots, [\delta_p\delta]$$

transform a into one and the same element $a_2 \neq a$. Moreover, these are the only elements of Δ which transform a into a_2. Similarly, there are just p elements in Δ which transform a into a_3, an element $\neq a, a_2; \cdots$; and there are just p elements which transform a into a_m, an element $\neq a, a_2, a_3, \cdots, a_{m-1}$. Thus $q = mp$, $m = q/p$, as was to be proved.

7. *Conjugate subhypergroups.* If H is an nGM and K a subhypergroup of H, then by Theorem 10 if the identity e of H is in K, K is an nGM whose nucleus Σ is a subgroup of the nucleus Δ of H.

Let K be any subhypergroup of an nGM H. Let δ be an element of the nucleus Δ of H. Then

$$[\delta^{-1}K\delta] = K_2$$

is a sub-nG of H conjugate to K. The set of all the conjugates of K is a complete set of conjugate subhypergroups. Let Δ_1 be the subgroup of all elements of Δ which are commutative with K; and suppose Δ is of finite order q and Δ_1 of order p. Then by an argument similar to that used in proving Theorem 16 we find that *the number of subhypergroups in the complete set of conjugates to which K belongs is q/p.*

In case the identity of H is in K we have this theorem.

THEOREM 17. *The elements of the nucleus Δ of an nGM H which are commutative with a sub-nG K, containing the identity of H, form a subgroup Δ_1 of Δ (with respect to the bracket product) which contains the nucleus Σ of K as an invariant subgroup.*

Proof. By Theorem 10, Σ is a subgroup of Δ. Let σ be any element of Σ; and let

$$K = \Sigma + [\Sigma a_2] + [\Sigma a_3] + \cdots$$

be a decomposition of K into mutually exclusive sets. Then

$$[\sigma^{-1}K\sigma] = \Sigma + [\Sigma a_2\sigma] + [\Sigma a_3\sigma] + \cdots = K.$$

Hence Σ is a subgroup of Δ_1. Moreover, if $\delta_1 \varepsilon \Delta_1$, then since

$$[\delta_1{}^{-1}K\delta_1] = K,$$

it must follow that $[\delta_1{}^{-1}\Sigma\delta_1] = \Sigma$, so that Σ is an invariant subgroup of Δ_1.

THEOREM 18. *If δ is a scalar and $[\delta^{-1}K\delta] = K_2$ is a sub-nG conjugate to the sub-nG K of H, then the group of elements of the nucleus Δ of H which are commutative with K_2 is $[\delta^{-1}\Delta_1\delta]$ where Δ_1 is the group of elements of Δ commutative with K.*

Proof. If $[\delta^{-1}\delta_1\delta] \varepsilon [\delta^{-1}\Delta_1\delta]$, then

$$\left[[\delta^{-1}\delta_1\delta]^{-1}K_2[\delta^{-1}\delta_1\delta]\right] = [\delta^{-1}\delta_1{}^{-1}\delta\delta^{-1}K\delta\delta^{-1}\delta_1\delta] = [\delta^{-1}K\delta] = K_2,$$

so that this element is commutative with K_2. Conversely, if $[d^{-1}K_2d] = K_2$, $d \varepsilon \Delta$, then

$$[d^{-1}\delta^{-1}K\delta d] = [\delta^{-1}K\delta], \qquad [[\delta d\delta^{-1}]^{-1}K[\delta d\delta^{-1}]] = K,$$

and therefore $[\delta d\delta^{-1}]$ is an element δ_1 of Δ_1. Thus $d = [\delta^{-1}\delta_1\delta] \,\varepsilon\, [\delta^{-1}\Delta_1\delta]$, as was to be proved.

THEOREM 19. *No sub-nGM of an nGM with nucleus of finite order whose nucleus is a proper subgroup of the nucleus of the main nG can contain elements belonging to every one of the complete sets of conjugates.*

Proof. Let H be the nGM, and suppose that such a sub-nG, K, can exist. Then the complete set of conjugate subhypergroups to which K belongs would contain all the elements of H. Let Δ_1 be the group of elements of Δ, the nucleus of H, commutative with K; and let Σ be the nucleus of K. Then the order v of Δ_1 is $\geq$ the order w of Σ by Theorem 17. Let N be the order of Δ. Then there are in all N/v subhypergroups in the conjugate set to which K belongs, and each of these contains w elements in its nucleus. Hence all these nuclii together cannot have more than

$$1 + w(N/v) - (N/v) = 1 + \frac{N(w-1)}{v} = 1 + \frac{Nw - N}{v} < 1 + \frac{Nw - w}{v} \leqq N$$

elements, which is impossible inasmuch as these nuclii together must make up Δ.

THEOREM 20. *The intersection of two or more sub-nG_1's of an nG_1 H, having a common identity element, is a sub-nG_1 of H.*

The proof of Theorem 20 will be omitted.

THEOREM 21. *If H is an nG_1M with finite or infinite nucleus, and K is a sub-nG_1 of H containing the identity of H, then the intersection of the complete set of conjugates to which K belongs is a normal sub-nG_1 of H i. e. is commutative with every element of the nucleus of H.*

Proof. Since K is an nG_1 containing the identity of H, it is easy to see that every one of the conjugates of K is an nG_1, and these have a common identity, namely the identity of H. Consequently, by Theorem 20, the intersection of these conjugates is a sub-nG_1 of H.

Denote by S the complete set of conjugates of K. Then since the nucleus Δ of H is a group, $[\delta^{-1}S\delta] = S$ for every $\delta \,\varepsilon\, \Delta$. Let T be the intersection of the subhypergroups in S. Then clearly $[\delta^{-1}T\delta] = T$, as was to be proved.

The following theorems are easy to verify.

THEOREM 22. *Let S be a complete set of conjugates in an nGM H. If*

S generates an nGM K, then K is a normal subhypergroup of H, and every other normal subhypergroup which contains an element of S must contain K as a subhypergroup.

THEOREM 23. *Let S be a complete set of conjugate subhypergroups in an nGM H. Then if these generate a subhypergroup K, the latter is a normal subhypergroup of H, and every other normal subhypergroup which contains one of the hypergroups of S must contain K.*

8. *Isomorphism of two hypergroups.* Two hypergroups H and H' of dimension n are *simply isomorphic* if the elements of H and H' can be put in one to one correspondence in such a way that if a and b are two elements of H which correspond to elements a' and b', respectively, of H', then the elements of the complexes ab and $a'b'$ are in one to one correspondence.

If H and H' are of type M with nuclii Δ and Δ', respectively, then Δ and Δ' are simply isomorphic groups. For, in the first place, the identity elements are in correspondence. In fact, if b' in Δ' corresponds to the identity element e in Δ, then

$$[ee] = e, \qquad [b'b'] = b'.$$

Consequently, by Theorem 6, $b' = e'$, the identity element of H'. If a' in H' corresponds to δ in Δ, then let a'^{-1} in H' correspond to δ^{-1}. Then

$$[\delta\delta^{-1}] = e \text{ implies that } [a'a'^{-1}] = e',$$
$$[\delta^{-1}\delta] = e \text{ implies that } [a'^{-1}a'] = e'.$$

Hence by Theorem 7 a' is a scalar and is therefore in Δ'.

Let a in H correspond to a' in H'. Then the complexes $[\Delta a]$ and $[\Delta' a']$ contain elements in one to one correspondence. We have proved this when a is in Δ. Hence suppose a is not in Δ. If $[\delta a]$ is an element of $[\Delta a]$ let a'_1 be the corresponding element in H'. Let δ^{-1} in Δ correspond to δ'^{-1} in Δ'. Then

$$[\delta^{-1}\delta a] = a, \qquad [\delta'^{-1}a'_1] = a', \qquad a'_1 = [\delta' a'],$$

so that a'_1 is in $[\Delta' a']$. We have this result.

THEOREM 24. *If H and H′ are simply isomorphic nGM's, and a in H corresponds to a′ in H′, then the elements of the complexes* $[\Delta a]$ *and* $[\Delta' a']$ *are in one to one correspondence, where* Δ *and* Δ' *are the nuclii of H and H′, respectively. The latter are simply isomorphic groups with respect to the bracket product.*

9. *Automorphism.* If $d = \begin{pmatrix} a \\ a' \end{pmatrix}$ is any substitution on the elements of

an $nG\,H$ which sets up a simple isomorphism of H with itself, we shall call d an *automorphism* of H.

THEOREM 25. *The set of all the automorphisms of a hypergroup forms a group.*

Proof. Clearly the identity substitution $\begin{pmatrix} a \\ a \end{pmatrix}$ is an automorphism; and if $d = \begin{pmatrix} a \\ a' \end{pmatrix}$ is an automorphism, then $d^{-1} = \begin{pmatrix} a' \\ a \end{pmatrix}$ is also an automorphism. Let $d_1 = \begin{pmatrix} a \\ a' \end{pmatrix}$ and $d_2 = \begin{pmatrix} a' \\ a'' \end{pmatrix}$ be two automorphisms. Then $d_1 d_2 = \begin{pmatrix} a \\ a'' \end{pmatrix}$.

But since d_1 is an automorphism, the relation

$$a_i a_j = c_1 + c_2 + \cdots + c_n \tag{9}$$

requires that

$$a'_i a'_j = c'_1 + c'_2 + \cdots + c'_n; \tag{10}$$

and since d_2 is an automorphism, the relation (10) requires that

$$a_i'' a_j'' = c_1'' + c_2'' + \cdots + c_n''. \tag{11}$$

Thus the relation (9) implies the relation (11), so that $d_1 d_2$ is an automorphism.

If H is an nGM with nucleus Δ, the substitution

$$\delta' = \begin{pmatrix} a \\ [\delta^{-1} a \delta] \end{pmatrix}, \ \delta \text{ in } \Delta,$$

is an automorphism of H called an inner automorphism of H.

THEOREM 26. *The set of inner automorphisms of an nGM, H, forms a group which is homomorphic with the nucleus of H. This group is an invariant subgroup of the group of automorphisms of H.*

Proof. Let $\delta'_i = \begin{pmatrix} a \\ [\delta_i^{-1} a \delta_i] \end{pmatrix}$ correspond to δ_i in Δ, and let $\delta_i \delta_j = \delta_k$. Then

$$\delta'_i \delta'_j = \begin{pmatrix} a \\ [\delta_i^{-1} a \delta_i] \end{pmatrix} \begin{pmatrix} a \\ [\delta_j^{-1} a \delta_j] \end{pmatrix}$$
$$= \begin{pmatrix} a \\ [\delta_i^{-1} a \delta_i] \end{pmatrix} \begin{pmatrix} [\delta_i^{-1} a \delta_i] \\ [\delta_j^{-1} \delta_i^{-1} a \delta_i \delta_j] \end{pmatrix} = \begin{pmatrix} a \\ [\delta_k^{-1} a \delta_k] \end{pmatrix} = \delta'_k.$$

Thus the inner automorphisms of H form a group homomorphic with Δ.

Now let

$$\begin{pmatrix} a \\ a' \end{pmatrix}$$

denote any automorphism of H. Then

$$\begin{pmatrix} a \\ a' \end{pmatrix}^{-1} \begin{pmatrix} a \\ [\delta^{-1}a\delta] \end{pmatrix} \begin{pmatrix} a \\ a' \end{pmatrix} = \begin{pmatrix} a \\ a' \end{pmatrix} \begin{pmatrix} a \\ [\delta^{-1}a\delta] \end{pmatrix} \begin{pmatrix} a \\ a' \end{pmatrix}$$

$$= \begin{pmatrix} a' \\ [\delta^{-1}a\delta] \end{pmatrix} \begin{pmatrix} a \\ a' \end{pmatrix} = \begin{pmatrix} a' \\ [\delta^{-1}a\delta] \end{pmatrix} \begin{pmatrix} [\delta^{-1}a\delta] \\ [\delta'^{-1}a'\delta'] \end{pmatrix} = \begin{pmatrix} a' \\ [\delta'^{-1}a'\delta'] \end{pmatrix}.$$

Thus the transform of any inner automorphism of H by any automorphism of H is also an inner automorphism of H. The group of inner automorphisms is an invariant subgroup of the group of automorphisms.

PART III.

Examples of Hypergroups.

10. *Three special examples.* Simple examples of 2-groups are as follows.

Example 1. The integers 1, -1, 2, -2 if the rule of combination is ordinary multiplication followed by a partition of the product into the sum of two elements of the set. This is a $2G_1M$.

Example 2. The numbers cos 0, cos 1, cos 2, $\cdots$ if the product is defined as twice the ordinary product according to the formula

$$(\cos m)(\text{con } n) = \cos(m+n) + \cos(m-n).$$

This is a $2G_1M$ of infinite order. The identity element is cos 0, the only scalar. The inverse of $\cos n$ is $\cos n$.

Example 3. The system defined by this multiplication table

	e	a	b
e	$e+e$	$a+a$	$b+b$
a	$a+a$	$a+b$	$e+b$
b	$b+b$	$e+b$	$a+a.$

Two other examples were given in § 3.

11. *Double cosets.* Let A be an ordinary group of which S and T are subgroups; and let $(A; S, T)$ denote the set of all the distinct double cosets SaT where a is in A. It is clear that SaT is completely determined by any one of its members. Thus if b is in SaT, then $SaT = SbT$.

We shall define the product of two elements SaT and SbT of $(A; S, T)$ by the formula [4]

[4] If S, T are not both finite this product contains an infinite number of terms. We shall suppose henceforth that S and T are finite groups.

$$(SaT)(SbT) = \Sigma(SatsbT)$$

where the summation runs independently over the elements t of T and s of S. The system so defined is a hypergroup. If 1 is the identity element in A, then $(S1T)$ is an identity of the hypergroup. An inverse of (SaT) is $(Sa^{-1}T)$, where a^{-1} is the inverse of a in A. The associative law holds since it holds in A.

If H is a hypergroup, then we shall define the transpose, H', of H as follows. Let h of H correspond to h' of H' in such a way that when

$$h_1h_2 = a_1 + a_2 + \cdots + a_n, \qquad h_2h_1 = b_1 + b_2 + \cdots + b_n,$$

then

$$h'_1h'_2 = b'_1 + b'_2 + \cdots + b'_n, \qquad h'_2h'_1 = a'_1 + a'_2 + \cdots + a'_n.$$

THEOREM 27. *The hypergroup $(A; T, S)$ is the transpose of the hypergroup $(A; S, T)$.*

Proof. If b is any element of (SaT), then b^{-1} is in $(Ta^{-1}S)$, and if b is in $(Ta^{-1}S)$ then b^{-1} is in (SaT). We let (SaT) of $(A; S, T)$ correspond to $(Ta^{-1}S)$ of $(A; T, S)$. The theorem now follows at once.

If $S = T = 1$, the identity of A, then $(A; S, T)$ reduces to the ordinary group A. If $T = 1$, $S \neq 1$, then $(A; S, T) = (A; S, 1)$ is simply isomorphic with the Mischgruppe of Loewy and Baer, insofar as multiplication is defined for the Mischgruppe, provided we use the bracket product.

Let A be a substitution group on the symbols $1, 2, 3, \cdots, k$. In (SaT) let there be p_{ij} substitutions which carry i into j, and form the matrix (p_{ij}) associated with (SaT). It is easy to prove the following theorem.

THEOREM 28. *If A is a substitution group on the symbols $1, 2, 3, \cdots, k$; and $\alpha = (SaT)$, $\beta = (SbT)$ are two elements of $(A; S, T)$ with product*

$$\alpha\beta = \gamma_1 + \gamma_2 + \cdots + \gamma_n$$

where $\gamma_i = (Sa_iT)$; and $\alpha', \beta', \gamma'_1, \gamma'_2, \cdots, \gamma'_n$ are the matrices associated with $\alpha, \beta, \gamma_1, \gamma_2, \cdots, \gamma_n$, respectively, then

$$\alpha'\beta' = \gamma'_1 + \gamma'_2 + \cdots + \gamma'_n.$$

The matrix associated with (SaT) is the transpose of the matrix associated with $(Ta^{-1}S)$.

Two different double cosets may have one and the same associated matrix. For example, the double cosets

$$(15)(24),\ (1542),\ (14235),\ (2354),\ (13425),\ (13)(254)$$

and

$$(254),\ (1425),\ (15423),\ (15)(234),\ (13542),\ (135)(24)$$

both have the associated matrix

$$\begin{pmatrix} 1 & 0 & 2 & 1 & 2 \\ 1 & 0 & 2 & 1 & 2 \\ 1 & 0 & 2 & 1 & 2 \\ 0 & 6 & 0 & 0 & 0 \\ 3 & 0 & 0 & 3 & 0 \end{pmatrix}.$$

In this example A is the symmetric group of degree 5, $S = \{(123)\}$, $T = \{(14)\}$. The two double cosets are $S(15)(24)T$ and $S(254)T$, respectively. This hypergroup contains no scalars and is of dimension 6, and order 20.

If two elements a and b of a hypergroup H have the property that $ax = bx$ $(xa = xb)$ for every x in H while $xa \neq xb$ $(ax \neq bx)$ for at least one x in H, then a and b are *semi-equal* elements. In the above example $a = S(24)T$ and $b = S(34)T$ are semi-equal: $ax = bx$, $xa \neq xb$ for every x.

In an ordinary group if three elements a, b, x are selected, an element y can be found such that $ax = by$. This is not in general true of a hypergroup. In the above example if $a = S(45)T$, $b = S(12)(45)T$, then for every x there is a y such that $ax = by$. Another exceptional pair of elements having this property is ST and $S(12)T$. For semi-equal elements $y = x$.

12. *Properties of the hypergroup* $(A; S, T)$. Let $B = \{S, T\}$ be the group generated by S and T, and let C be the group of all elements of A which are commutative with B[5]. If $a \varepsilon B$ then every element in the double coset SaT is in B. Let $(A; S, T)_0$ be the subset of elements of $(A; S, T)$ which are double cosets made up of elements of B.

THEOREM 29. $(A; S, T)_0 = (B; S, T)$ *is a sub-hypergroup of* $(A; S, T)$ *which has the following properties. The product of an element in* $(A; S, T)_0$ *and an element of* $(A; S, T)$ *not in* $(A; S, T)_0$ *is a complex of elements not in* $(A; S, T)_0$; $(A; S, T)$ *can have no identity element outside* $(A; S, T)_0$: *no element in* $(A; S, T)_0$ *can have an inverse outside* $(A; S, T)_0$.

Proof. The last two statements follow from the first. Let (SaT) be in $(A; S, T)_0$ and (SbT) not in $(A; S, T)_0$. Then

$$(SaT)(SbT) = \Sigma(SatsbT).$$

[5] C is the "normalizer" of B in A.

Now if $atsb = c$ is in B, then $b = a^{-1}t^{-1}s^{-1}c$ is in B, contrary to hypothesis. Likewise the complex $(SbT)(SaT)$ contains no elements in $(A; S, T)_0$.

If a is in C then every element of (SaT) is in C. Let $(A; S, T)_1$ be the subset of elements of $(A; S, T)$ which are double cosets made up of elements of C. The following theorem is evident.

THEOREM 30. *$(A; S, T)_0$ is a sub-hypergroup of $(A; S, T)_1$; and $(A; S, T)_1 = (C; S, T)$ is a sub-hypergroup of $(A; S, T)$. The product of an element in $(A; S, T)_1$ and an element outside $(A; S, T)_1$ is a complex of elements outside $(A; S, T)_1$; no element in $(A; S, T)_1$ can have an inverse outside $(A; S, T)_1$.*

It will be seen that if U is the group of all elements common to S and T, then each element of $(A; S, T)$ is a sum of elements of $(A; U, U)$; and that each element of $(A; B, B)$ is a sum of elements of $(A; S, T)$.

Let p be any element of A, $S' = p^{-1}Sp$, $T' = p^{-1}Tp$. Let $a' = p^{-1}ap$, where a is in A. Then we have

$$(SaT)(SbT) = \Sigma(SatsbT), \qquad (S'a'T')(S'b'T') = \Sigma(S'[atsb]'T').$$

We therefore have this result.

THEOREM 31. *The hypergroups $(A; S, T)$ and $(A; p^{-1}Sp, p^{-1}Tp)$, where p is in A, are simply isomorphic hypergroups.*

The totality of all the elements of the hypergroups $(A; S_i, T_j)$, $i, j = 1, 2, 3, \cdots$, where S_i, T_i are conjugates of S_1, T_1, respectively, in A, forms a system satisfying postulates I, II § 1, if we define the product of two elements by the formula

$$(S_iaT_j)(S_kbT_l) = \Sigma(S_iat_js_kbT_l),$$

where the summation is taken over the elements t_j of T_j and s_k of S_k.

13. *Properties of the hypergroup $(A; U, U)$.* We have seen that every hypergroup $(A; S, T)$ is a set of complexes of elements of a hypergroup $(A; U, U)$ where U is the set of elements common to S and T.

THEOREM 32. *The hypergroup $(A; U, U)$ is of type M and every element has a unique inverse.*

Proof. The element $(U1U)$ is evidently a scalar, so that $(A; U, U)$ is of type M. If $(U1U) \varepsilon (UaU)(UbU)$ then there exists an element $u' \varepsilon U$ such that

$$au'b = 1, \qquad b = u'^{-1}a^{-1}, \qquad (UbU) = (Ua^{-1}U).$$

Hence every element has a unique inverse.

In the product of two elements of $(A; U, U)$ every element is repeated n times, where n is the order of U; and

$$(UaU)(UbU) = n\Sigma(UaubU)$$

where the summation is taken over the elements u of U. The system $(A; U, U)'$ made up of the elements of $(A; U, U)$ but with the product definition

$$(UaU)(UbU) = \Sigma(UaubU)$$

is a hypergroup of dimension n.

Example. Let U be of order 2. Then $(A; U, U)'$ is a 2-group. Let

$$(UaU)(Ua^{-1}U) = (U1U) + (UcU).$$

Clearly $c^2 = 1$, so that (UcU) is its own inverse. This shows that the hypergroup of Example 3, § 10, cannot be represented as a hypergroup $(A; U, U)'$, where U is of order 2.

The matrix associated (see § 11) with (UaU) is the transpose of the matrix associated with $(Ua^{-1}U)$.

14. *Cyclic hypergroups.* If a hypergroup H is generated by a single element a of H, then H will be called a *cyclic hypergroup.* In a cyclic hypergroup the commutative law does not necessarily hold. We may write

$$aa = a^2, aaa = a^3, \cdots;$$

and clearly $a^m a^n = a^{m+n}$. It is evident that the set of elements generated by a single element of a hypergroup is closed, i. e. satisfies postulate I, § 1, but I am unable to say whether or not it forms a hypergroup.

In connection with a cyclic nG one would naturally seek an analog of the representation of a cyclic group as the group of rotations of a regular polygon. This leads to the notion of cyclomorphism. Let H be any nG, and let $d = \begin{pmatrix} a \\ a' \end{pmatrix}$

be a substitution on the elements of H such that $ka = \Sigma c_i$ implies that $ka' = \Sigma c'_i$ where $a \sim a'$ and $c_i \sim c'_i$ $(i = 1, 2, 3, \cdots, n)$, for every a in H, and k is an assigned element of H. Then d will be called a *cyclomorphism* of H relative to k. The set $G(k)$ of all cyclomorphisms of H relative to k is clearly a group.

If H is cyclic with generator k we may represent it in a certain sense as a system of points and directed lines. Let each element of H be identified with a certain point in a space S. From the point k draw directed lines to each of the points in the set $[k^2]$ which are different from k. Call them $k_1, k_2, \cdots, k_p$. If k_i occurs m_i times in k^2 we shall consider the line from k to k_i as an m_i-tuple line. Then from each point k_i draw lines to the elements in $[kk_i]$ which are different from k_i, etc. The group $G(k)$ of cyclomorphisms of H relative to k is a group of transformations of this configuration of points and lines into itself.

As an illustration let $H = (A; U, U)'$ where A is the symmetric group of degree 5, and $U = \{(12)\}$. Then H is of order 33 with the elements

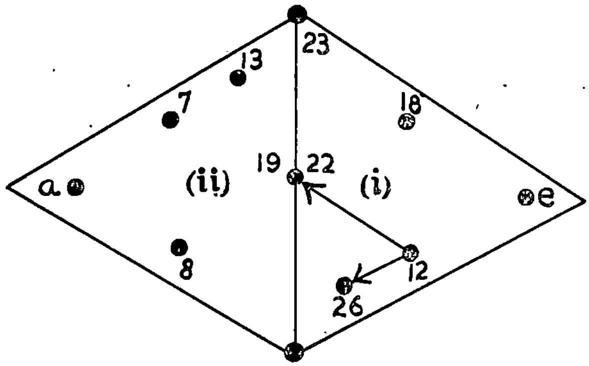
$$e = U1U, \quad a = U(34)U, \quad b = U(35)U, \quad c = U(45)U, \quad d = U(345)U,$$
$$f = U(354)U$$

which are scalars; and the non-scalars

$$1 = U(13)U, \ 2 = U(14)U, \ 3 = U(15)U, \ 4 = U(134)U, \ 5 = U(135)U,$$
$$6 = U(143)U, 7 = U(145)U, 8 = U(153)U, 9 = U(154)U, 10 = U(1324)U,$$
$$11 = U(1325)U, \ 12 = U(1345)U, \ 13 = U(1354)U, \ 14 = U(1425)U,$$
$$15 = U(1435)U, \ 16 = U(1453)U, \ 17 = U(1534)U, \ 18 = U(1543)U,$$
$$19 = U(13245)U, \ 20 = U(13254)U, \ 21 = U(13425)U, \ 22 = U(13524)U,$$
$$23 = U(14253)U, 24 = U(14325)U, 25 = U(13)(45)U, 26 = U(14)(35)U,$$
$$27 = U(15)(34)U.$$

This $2G_1M$ may be generated by $k = 12$; and $G(k)$ is simply isomorphic with the group of rotations of a six sided double pyramid into itself. This suggests that we represent the elements as points upon this double pyramid.

Hold the solid with the principal axis vertical and, looking downward, number the faces of the upper tetrahedron i, iv, iii counterclockwise. Number the faces below these on the lower tetrahedron ii, v, vi, respectively. On face i mark the elements 23, 18, *e*, 12, 26, 22 as shown in the accompanying figure.



The faces iv and iii have these elements replaced by 20, 1, *d*, 15, 9, 14 and 10, 16, *f*, 3, 4, 21, respectively. The face numbered ii is marked with the elements 8, *a*, 7, 13, 19 as shown. The faces v and vi have these elements replaced by 25, *b*, 27, 2, 24 and 6, *c*, 5, 17, 11, respectively. The points representing 19 and 22, 14 and 24, 11 and 21 are considered as double points. Now $12 \cdot 12 = 26 + 22$. Draw arrows from 12 to 26 and to 22. Then $12 \cdot 26 = 18 + 24$. Draw arrows from 26 to 18 and to 24. Likewise draw arrows from 22 to 23 and to 20, and so on until the configuration is complete. The six rotations of the solid into itself carry the configuration of points and lines into itself.

Added in proof. Since this paper was written a paper by M. F. Marty: "Sur les groupes et hypergroupes," *Annales de l'École Normale*, (3), vol. 53, pp. 83-123, has appeared. Marty applies the special hypergroup $[A; U, U]$ of § 13 to a problem in topology. His results do not overlap mine.

NORTHWESTERN UNIVERSITY,
EVANSTON, ILL.

PRIMARY ABELIAN GROUPS AND THEIR AUTOMORPHISMS.[1]

By Reinhold Baer.

The object of this paper is to investigate the relations between automorphisms and properties of the group itself. To every group of automorphisms corresponds first the group of its fixed elements and secondly the subgroup, generated by the "commutators" $x - xf$ of the elements x of the group under the automorphisms f of the given group of automorphisms. This commutator group of a group of automorphisms A is exactly the smallest subgroup S such that A induces mod S the identical automorphism. Those subgroups of primary Abelian groups which are complete groups of fixed elements and those which are complete commutator groups of suitable groups of automorphisms are characterized by their structure and their situation in the whole group.

Furthermore those automorphisms are studied which map every subgroup upon itself. It is proved that the centralizer of the group of all the automorphisms is essentially formed by these automorphisms.

Finally some instances of the problem to characterize a group by properties of its automorphisms are considered. In this respect the following two classes of primary Abelian groups G are enumerated:

The group of all the automorphisms of G is Abelian.

Every automorphism of the lattice of the subgroups of G is induced by an automorphism of G.

The proofs are based on certain decomposition-theorems for primary Abelian groups which are enunciated and proved in section 1.

1. Invariants and decompositions of primary Abelian groups.[2]

If G is an (additively written primary Abelian) group (of characteristic p), then the *order* of every element x in G is a power $p^{n(x)}$ of the prime number p. The elements x in G with $n(x) \leq i$ form a characteristic subgroup $(G, n(x) \leq i)$ of G.

The *rank* $r(G)$ of the group G is the number of elements in $(G, n(x) \leq 1)$, if $(G, n(x) \leq 1)$ is infinite, and if $(G, n(x) \leq 1)$ is finite, then $(G, n(x) \leq 1)$

[1] Presented to the American Mathematical Society, October 31, 1936. Received by the Editors September 22, 1936.

[2] In this section, those concepts and facts of the theory of primary Abelian groups are collected which are needed in the course of this paper.

contains exactly $p^{r(G)}$ elements. Thus G contains $r(G)$ elements, if $r(G)$ is infinite.

Since the orders of the elements of $(G, n(x) \leq i)$ are bounded, every $(G, n(x) \leq i)$ with positive i is a direct sum of $r(G)$ cyclic groups.[3]

Denote by $p^{\omega}G$ the intersection of all the groups p^iG (for integral i)[4] and by $p^{\infty}G$ the uniquely determined greatest subgroup S of G, satisfying $S = pS$. Every subgroup S of G which satisfies $S = pS$ is a direct summand of G and a direct sum of groups of type p^{∞}.[5,6]

Put $r(G, i) = r((p^{i-1}G, n(x) \leq 1)/(p^iG, n(x) \leq 1))$ for integral i and $r(G, \infty) = r(p^{\infty}G)$. Then the structure of G is completely determined by these invariants, provided G is a direct sum of cyclic groups and of groups of type p^{∞}. If in particular G is the direct sum of groups G_v, i. e.

$$G = \sum_v G_v,$$

then at most $r(G, i)$ of the groups G_v are cyclic groups of order p^i and at most $r(G, \infty)$ of the groups G_v are groups of type p^{∞}.

Conversely: If m is a positive integer, then

$$G = \sum_{i=1}^{m} G_i + p^{\infty}G + G',$$

where G_i is a direct sum of $r(G, i)$ cyclic groups of order p^i, $p^{\infty}G$ a direct sum of $r(G, \infty)$ groups of type p^{∞} and

$$r(G', 1) = \cdots = r(G', m) = r(G', \infty) = 0.$$

A proof of this theorem[7] may be given: $G = G'' + p^{\infty}G$, as has been mentioned before, and $r(G'', i) = r(G, i)$ for every finite i, $r(G'', \infty) = 0$. There exists for every integer i a subset B_i of G'' such that the elements $p^{i-1}b$ with b in B_i form a basis of $(p^{i-1}G'', n(x) \leq 1)$ mod $(p^iG'', n(x) \leq 1)$ (and of $(p^{i-1}G, n(x) \leq 1)$ mod $(p^iG, n(x) \leq 1)$). The subgroup G_i of G'', generated by B_i, is then a direct sum of $r(G, i)$ cyclic groups of order p^i.

Denote by M the subgroup of G'', generated by the elements in sets B_i

---

[3] A proof of this well known theorem may be found e. g. in R. Baer, "Der Kern, eine characteristische Untergruppe," *Comp. Math.*, vol. 1 (1934), pp. 254-283, Lemma in § 5.

[4] p^iG consists of all the elements p^ix for x in G and it is obvious how to define p^vG for every ordinal number v by complete transfinite induction.

[5] A group of type p^{∞} is generated by elements g_i, satisfying $n(g_0) = 1$, $pg_i = g_{i-1}$.

[6] Cf. e. g. L. Zippin, "Countable torsion groups," *Annals of Mathematics*, vol. 36 (1935), pp. 86-99.

[7] Similar results and arguments may be found in the investigations of H. Prüfer.

with $m < i$, and let G' be the set of all elements x in G'' such that there exists an element x' in M, satisfying $x \equiv x' \bmod pG''$. Then G' is a subgroup of G'', G is the direct sum of the groups $G_1, \cdots, G_m, p^\infty G, G'$ and this direct decomposition of G meets the requirements of the above proposition.

It is an important corollary, derived from the above proof, that the set B_i is a basis of G_i and may be chosen at random, only restricted by the condition that the elements $p^{i-1}b$ with b in B_i form a basis of

$$(p^{i-1}G, n(x) \leq 1) \bmod (p^iG, n(x) \leq 1).$$

If N is the subgroup of G, generated by the elements in all the sets B_i of the above construction, then $G/N = p(G/N)$. If the orders of the elements in G are not bounded, then it is possible to choose the sets B_i in such a way that $G/N \neq 0$. This implies the following theorem:

G is homomorphic to a group of type p^∞ if, and only if, the orders of the elements in G are not bounded.[8]

2. Automorphisms.

Every single-valued additive function $\boldsymbol{f}$ of the elements of the (additively written Abelian) group G with values $x\boldsymbol{f}$ in G is an automorphism of G. The set $\boldsymbol{R}(G)$ of all the automorphisms of G is a ring under the well known definition of addition and multiplication of automorphisms.

The proper automorphisms of G (which define a one-one correspondence between all the elements of G such that an inverse automorphism exists) form a multiplicative group $\boldsymbol{P}(G)$. The identity element of $\boldsymbol{P}(G)$ may be denoted by $\mathbf{1}$.

If $\boldsymbol{S}$ is any set of automorphisms of G; $g(x) = \sum_{i=0}^{n} c_i x^i$ a polynomial with integer coefficients, S and T subsets of G, then

$(S, xg(\boldsymbol{S}) \leq T)$ is the set of all the elements x in S which are mapped by all the automorphism $g(\boldsymbol{s})$ with $\boldsymbol{s}$ in $\boldsymbol{S}$ upon elements in T; and

$(\boldsymbol{S}, Sg(\boldsymbol{x}) \leq T)$ is the set of all the automorphisms $\boldsymbol{x}$ in $\boldsymbol{S}$ such that $g(\boldsymbol{x})$ maps S upon a subset of T.[9]

Some particular instances of the correspondence[10] between sets of auto-

---

[8] Cf. R. Baer, "Dualisms in Abelian groups," *Bulletin of the American Mathematical Society.*

[9] Also $(S, T \leq xg(\boldsymbol{S}))$ and $(\boldsymbol{S}, T \leq Sg(\boldsymbol{x}))$ may be considered.

[10] Such a correspondence has first been introduced by K. Shoda, "Über den Automorphismenring bezw. die Automorphismengruppe einer endlichen Abelschen Gruppe," *Proceedings of the Imperial Academy of Japan*, vol. 6 (1930), p. 10.

morphisms and sets of elements, thus defined, shall be investigated in the course of this paper.

It is easily verified, that

to every subgroup S of G and to every isomorphism g of G/S upon a subgroup T of G there exists one and only one automorphism f of G such that $S = (G, xf = 0)$ and g is the isomorphism of G/S upon $T = Gf$, induced by f. Conversely every automorphism f induces an isomorphism of $G/(G, xf = 0)$ upon Gf.

Since the automorphisms f and $1 - f$ determine the one the other completely, the above result implies, that the automorphism f of G is completely determined by the group $(G, xf = x)$ of the fixed elements and the induced isomorphism of $G/(G, xf = x)$ upon $G(1 - f)$.

3. Fixed elements.

Notation. If S is a subgroup of G, then $F(S)$ is the subgroup of all those elements of G which are mapped upon themselves by all the proper automorphisms of G which map every element of S upon itself, i. e.

$$F(S) = (G, x(\boldsymbol{P}(G), S(1 - f) = 0) = x).$$

Then $S \leq F(S)$ for every S.

Theorem 3.1.[11] *Suppose that G is a primary Abelian group of characteristic p, and that $p^{\omega}G = p^{\infty}G$. Then*

$$F(0) \neq 0$$

if, and only if,

(a) $p = 2$;

(b) *either* $r(G, \infty) = 1$ *or there exists an integer* m *such that*

$$r(G, m) = 1, \qquad p^m G = 0.$$

If the conditions (a) *and* (b) *are satisfied, then* $F(0)$ *is a group of order* 2 *and*

$$F(0) = (p^{\infty}G, n(x) \leq 1) \quad \text{or} \quad F(0) = (p^{m-1}G, n(x) \leq 1) = p^{m-1}G.$$

Proof. Since $xr = -x$ is a proper automorphism of G, $2x = 0$ for every element x in $F(0)$. Thus $F(0) \neq 0$ implies $p = 2$.

If g and h are different elements of order 2 in G, then

$$G = A + B + G',$$

[11] Cf. R. Baer, "Types of elements and characteristic subgroups of Abelian groups," *Proceedings of the London Mathematical Society*, ser. 2, vol. 39 (1935), pp. 481-514, in particular Theorem 10.

where g is the only element of order 2 in A, h the only element of order 2 in B, as has been pointed out in section 1. A and B are either cyclic groups or of type 2^∞ and it may therefore be assumed that A is isomorphic with a subgroup of B. Then there exists an automorphism of G which maps every element of $B + G'$ upon itself and g upon $g + h$, i. e. g does not belong to $F(0)$. The necessity of the conditions is now a consequence of the theorems in section 1.

The conditions are sufficient, since a group of order 2 contains but one element $\neq 0$, and since $(p^iG, n(x) \leq 1)$ and $(p^\infty G, n(x) \leq 1)$ is mapped upon itself by every proper automorphism of G.

LEMMA 3. 2. *Assume that S is a subgroup of the primary Abelian group G of characteristic p.*

(a) *If G/S is a cyclic group, then $F(S)$ is the join of S and $F(0)$.*

(b) *If G/S is a group of type p^∞, then*

$$\begin{aligned} F(S) &= G && \textit{for } p^\infty G = 0, \\ &= \textit{join of } S \textit{ and } F(0) && \textit{for } p^\infty G \neq 0. \end{aligned}$$

Proof. Note first that $S \leq F(S)$ and $F(0) \leq F(S)$.

If G/S is a cyclic group of order p^n with $0 < n$, then there exists an element g in G which generates G mod S. Put $g^* = p^n g$ and $m = n(g^*)$. Every element of G can be represented in one and only one way in the form

$$s + cg \text{ with } s \text{ in } S \text{ and } 0 \leq c < p^n.$$

If not at the same time $m = 0$ and $p = 2$, i. e. if $1 + p^m$ is relatively prime to p, then a proper automorphism f of G is defined by

$$(s + cg)f = s + c(1 + p^m)g$$

and

$$S = (G, xf = x).$$

If $m = 0$ and $p = 2$, then $g^* = 0$ and $G = \bar{g} + S$ where $\bar{g}$ is the cyclic group, generated by g, since $\bar{g}$ and S generate G and have only 0 in common. If S contains an element h of order 2^n, then an automorphism f of G is defined by

$$gf = g + h, \quad sf = s \text{ for } s \text{ in } S$$

and

$$S = (G, xf = x).$$

If S does not contain an element of order 2^n, then $r(G, n) = 1$, $2^nG = 0$, i. e. $F(0)$ is generated by $2^{n-1}g$. An automorphism f of G is defined by

$$gf = -g, \quad xf = x \text{ for } x \text{ in } S$$

and $$(G, xf = x) = \text{join of } S \text{ and } F(0).$$

Assume now that G/S is a group of type p^∞. If f is an automorphism of G such that $sf = s$ for every s in S, then $1 - f$ maps G and G/S upon a subgroup of G of type p^∞ or upon 0. Thus $p^\infty G = 0$ implies $G = F(S)$.

Assume now that $p^\infty G \neq 0$.

If there exists a subgroup Z of type p^∞ of G which contains elements $\neq 0$ of S, then denote by $z(i)$ a sequence of elements in Z such that $pz(i) = z(i-1)$ and $n(z(1)) = 1$. $z(1)$ is an element of S. There exists furthermore a sequence $g(i)$ in G such that $pg(i) \equiv g(i-1) \bmod S$ and $n(S + g(1)) = 1$. A proper automorphism f of G is defined by

$$sf = s \text{ for } s \text{ in } S, \qquad g(i)f = g(i) + z(i),$$

since $g(1)f \equiv g(1) \bmod S$,

$$pg(i)f = pg(i) + pz(i) \equiv g(i-1) + z(i-1) \equiv g(i-1)f \bmod S,$$

and since G/S is a group of type p^∞. Since $g(i)f \neq g(i)$, $S = (G, xf = x)$.

If none of the subgroups of type p^∞ of G contains an element $\neq 0$ of S, then $p^\infty G$ is a group of type p^∞, since G/S is of type p^∞, and consequently

$$G = S + p^\infty G.$$

Since by Theorem 3.1, $F(0) = (p^\infty G, n(x) \leq 1)$, the automorphism f, defined by $sf = s$ for s in S, $xf = -x$ for x in $p^\infty G$, satisfies

$$(G, xf = x) = \text{join of } S \text{ and } F(0).$$

Thus the proof of the Lemma 3.2 is complete. Note that in the discussed cases $F(S)$ is the group of fixed elements of a single proper automorphism.

THEOREM 3.3. *Assume that S is a subgroup of the primary Abelian group G of characteristic p.*

(a) *If $p^\infty G \neq 0$, then $F(S)$ is the join of S and $F(0)$.*

(b) *If $p^\omega G = 0$, then $F(S)$ is the join of $W(S)$ and $F(0)$ where $W(S)$ is the subgroup of G, satisfying $S \leq W(S)$ and $W(S)/S = p^\omega(G/S)$.*[12]

Proof. If f is any proper automorphism of G such that the elements of

[12] Since every characteristic subgroup $\neq 0$ of a direct sum of irreducible groups has the form, required in (a) and (b) respectively, Theorem 11 of the paper, mentioned in footnote [11] is a consequence of this Theorem.

S are fixed elements of f, then $1-f$ maps G/S homomorphic upon a subgroup of G and maps therefore $W(S)/S = p^\omega(G/S)$ upon a subgroup of $p^\omega G$. Thus $W(S) \leq F(S)$, if $p^\omega G = 0$. Since $p^\omega(G/W(S)) = 0$, it follows from the theorems of the section 1, that every element $W(S) + g$ of order p is contained in a cyclic direct summand of $G/W(S)$. There exists therefore for every element g of G which is not contained in $W(S)$ a subgroup H of G such that $W(S) \leq H$, G/H is a cyclic group and g is not contained in H. Then it follows from Lemma 3.2, (a) that $F(H)$ is the join of $F(0)$ and H and now it is easily seen that $F(S)$ is the join of $F(0)$ and $W(S)$, if $p^\omega G = 0$.

If $p^\infty G \neq 0$, then denote by S' the join of S and $F(0)$. If g is any element of G which is not contained in S', then there exists a greatest subgroup H of G, containing S' but not g. G/H is either a cyclic group or a group of type p^∞ and it follows from Lemma 3.2 that $H = F(H)$, since $F(0) \leq H$ and this completes the proof.

Since (as pointed out in section 1) G is not homomorphic to a group of type p^∞ if, and only if, the orders of the elements of G are bounded, Lemma 3.2 and Theorem 3.3 imply the

COROLLARY 3.4. *$F(S)$ is for every subgroup S of G the join of S and $F(0)$ if, and only if, either $p^\infty G \neq 0$ or the orders of the elements of G are bounded.*

COROLLARY 3.5. *The subgroup S of the primary Abelian group G with $p^\omega G = 0$ satisfies $F(S) = S$ if, and only if, $F(0) \leq S$ and $p^\omega(G/S) = 0$.*

4. Commutators.

NOTATIONS. If $\mathbf{S}$ is any group of (proper) automorphisms of the Abelian group G, then $(G, \mathbf{S})$ is the subgroup of G, generated by the elements $x - x\mathbf{s}$ with x in G and $\mathbf{s}$ in $\mathbf{S}$.

If S is any subgroup of G, then $(\mathbf{P}(G), S)$ is the set (and therefore the group) of all the proper automorphisms $\mathbf{f}$ of G such that $x - x\mathbf{f}$ is an element of S for every x in G.

$C(S) = (G, (\mathbf{P}(G), S))$ is a subgroup of G, contained in S.

$(G, \mathbf{S})$ is the cross cut of all the subgroups T of G such that $\mathbf{S}$ induces the identical automorphism in G/T.

$(G, \mathbf{S}) \leq (G, \mathbf{T})$, if $\mathbf{S} \leq \mathbf{T}$ and $(\mathbf{P}(G), S) \leq (\mathbf{P}(G), T)$, if $S \leq T$.
$(G, \mathbf{S}) = (G, (\mathbf{P}(G), (G, \mathbf{S})))$, $(\mathbf{P}(G), S) = (\mathbf{P}(G), (G, (\mathbf{P}(G), S)))$.
$(\mathbf{P}(G), G) = \mathbf{P}(G)$, $(\mathbf{P}(G), 0) = 1$.
$(G, 1) = 0$.

If $\boldsymbol{S}$ is a self conjugate subgroup of $\boldsymbol{P}(G)$, then $(G, \boldsymbol{S})$ is a characteristic subgroup of G, and if S is a characteristic subgroup of G, then $(\boldsymbol{P}(G), S)$ is a self conjugate subgroup of $\boldsymbol{P}(G)$.

THEOREM. *Assume that S is a subgroup of the primary Abelian group G of characteristic p.*

(a) *If the orders of the elements of $G/p^{\omega}G$ are not bounded, then $S = C(S)$.*

(b) *If p^m is the maximum order of the elements of $G/p^{\omega}G$ and either $p \neq 2$ or $p = 2$, but $1 < r(G, m)$, then $C(S)$ is the join of $p^{\infty}S$ and $(S, n(x) \leqq m)$.*

(c) *If $p = 2$, 2^m is the maximum order of the elements of $G/2^{\omega}G$ and $r(G, m) = 1$, then $C(S)$ is the join of $2^{\infty}S$, $(S, n(x) < m)$ and $(S; n(x) \leqq m, 2^{m-1}x$ in $2^{\omega}G)$.*

Proof. Note first that $C(S) \leqq S$ for every S.

If i is an integer such that $r(G, i) \neq 0$, then there exists a cyclic direct summand of order p^i in G, i. e.

$$G = Z + Z'$$

where Z is a cyclic group of order p^i, generated by z. If s is any element in S such that $n(s) < i$, then an automorphism f in $(\boldsymbol{P}(G), S)$ is defined by

$$xf = x \text{ for } x \text{ in } Z', \qquad zf = z + s$$

and s is an element of $C(S)$. Hence

$$(S, n(x) < i) \leqq C(S) \text{ for every } i \text{ with } r(G, i) \neq 0.$$

If the orders of the elements of $G/p^{\omega}G$ are not bounded, then there exists an infinity of integers i such that $r(G, i) \neq 0$, since $r(G, i) = r(G/p^{\omega}G, i)$ for every finite i. This provès (a), since S is the join of all the groups $(S, n(x) < i)$ with integral i.

If the orders of the elements of $G/p^{\omega}G$ are bounded, then there exists a maximum order p^m of the orders of the elements of $G/p^{\omega}G$ and $r(G, m) \neq 0$, $r(G, i) = 0$ for $m < i$. Then the above result shows that

$$(S, n(x) < m) \leqq C(S).$$

If $s \neq 0$ is an element of $p^{\infty}S$, then there exist elements $s(i)$ in $p^{\infty}S$ such that $s(0) = s$, $ps(i) = s(i-1)$ for $0 < i$. The elements $s(i)$ generate a direct summand D of S and G, i. e.

$$G = D + D'$$

and an automorphism f in $(\boldsymbol{P}(G), S)$ is defined by

$$xf = x \text{ for } x \text{ in } D', \qquad s(i)f = s(i) + s(i-1) \text{ for } 0 < i,$$

since $n(s(i-1)) < n(s(i))$ and

$ps(i)f = ps(i) + ps(i-1) = s(i-1) + s(i-2) = s(i-1)f$ for $1 < i$.

s is therefore an element of $C(S)$, i. e.

$$p^{\infty}S \leq C(S).$$

By section 1 there exists a direct decomposition

$$(*) \qquad G = \sum_{i=1}^{m} G_i + p^{\infty}G$$

of G where G_i is a direct sum of $r(G, i)$ cyclic groups of order p^i. If s is an element of order p^m in S, then

$$s = u + v + w$$

with u in $\sum_{i=1}^{m-1} G_i$, v in G_m and w in $p^{\infty}G$, $n(w) \leq m$.

If $n(v) < m$, i. e. if $p^{m-1}s$ is an element of $p^{\infty}G$, then let Z be a cyclic direct summand of order p^m of G, generated by z, i. e.

$$G = Z + Z'$$

and an automorphism f in $(\boldsymbol{P}(G), S)$ is defined by

$$xf = x \text{ for } x \text{ in } Z', \qquad zf = z + s.$$

Consequently s is an element of $C(S)$, i. e.

$$(S; n(x) \leq m, p^{m-1}x \text{ in } p^{\omega}G) \leq C(S).$$

If (in the above representation of the element s in S) v is an element of order p^m, then v generates a direct summand $\bar{v}$ of G, i. e.

$$G = \bar{v} + V$$

and, if $p \neq 2$, an automorphism f in $(\boldsymbol{P}(G), S)$ is defined by

$$xf = x \text{ for } x \text{ in } V, \qquad vf = v + s = 2v + u + w.$$

Consequently s is an element of $C(S)$ and

$$(S, n(x) \leq m) \leq C(S), \text{ if } p \neq 2.$$

If $1 < r(G, m)$ and $n(v) = m$, then

$$G = \bar{v} + \bar{z} + G'$$

where $\bar{z}$ is the cyclic group of order p^m, generated by z. An automorphism f in $(\boldsymbol{P}(G), S)$ is then defined by

$$xf = x \text{ for } x \text{ in } G' + \bar{v}, \qquad zf = z + s,$$

since $p^{m-1}s$ is an element of $\bar{v} + G'$. Thus s is an element of $C(S)$ and

$$(S, n(x) \leq m) \leq C(S), \quad \text{if} \quad 1 < r(G, m).$$

Let finally f be any automorphism in $(\boldsymbol{P}(G), S)$. Then $1 - f$ maps $\sum_{i=1}^{m-1} G_i$ (cf. (*)!) upon a subgroup of $(S, n(x) < m)$, $p^\infty G$ upon a subgroup of $p^\infty S$ and G_m upon a subgroup of $(S, n(x) \leq m)$ and that proves (b). If $p = 2$ and $r(G, m) = 1$, i. e. if G_m is a cyclic group of order 2^m, generated by g, then

$$gf - g = b + cg + d$$

with b in $\sum_{i=1}^{m-1} G_i \leq (G, n(x) < m)$ and d in $p^\infty G$. Since f is a proper automorphism of G, $1 + c$ is relatively prime to $p = 2$, i. e. c is even and $2^{m-1}(gf - g) = 2^{m-1}d$ is an element of $2^\infty G$, i. e. G_m is mapped by $1 - f$ upon a subgroup of $(S; n(x) \leq m, 2^{m-1}x \text{ in } 2^\omega G)$ and this completes the proof of (c).

If in particular $S = G$, then the theorem implies the

COROLLARY. *Assume that G is a primary Abelian group of characteristic p. Then $C(G) \neq G$ if, and only if,*

(1) $p = 2$;

(2) *there exists an integer m such that $r(G, m) = 1$, $2^m G = 2^\omega G$.*

If (1) and (2) are satisfied, then $C(G)$ is the join of $2^\omega G$ and $(G, n(x) < m)$, and $G/C(G)$ is of order 2.

5. Automorphisms of the lattice of the subgroups.

If $L(G)$ is the set of all the subgroups of the Abelian group G, then an

automorphism f of $L(G)$ is a single valued function of the elements of $L(G)$ with values in $L(G)$, satisfying:

$$L(G)f = L(G),$$

$$S \leq T \text{ if, and only if, } Sf \leq Tf,$$

S and Sf contain the same number of elements,
if $S < T$, then T/S and $(Tf)/(Sf)$ contain the same number of elements.[13]

Every automorphism of the lattice $L(G)$ of the subgroups of G is therefore a one-one correspondence between the subgroups of L, mapping 0 upon 0, G upon G.

Every automorphism of G induces an automorphism of $L(G)$, but the converse is generally not true, as the following theorem shows.

THEOREM 5.1. *Assume that G is a primary Abelian group of characteristic p. Then every automorphism of $L(G)$ is induced by an automorphism of G if, and only if, either G is irreducible, or G is a direct sum of a group of order 2 and of an irreducible group.*[14]

Proof. Since by section 1 irreducible primary groups are cyclic groups or groups of type p^∞, it is well known [15] that every automorphism of $L(G)$ is induced by an automorphism of G, if G is irreducible. If $G = T + W$ where T is of order 2 and W irreducible, then there exists to every cyclic subgroup Z of W such that $2W \neq 0$ exactly one subgroup Z' of G such that $Z' \neq Z$, $2Z' = 2Z$. If W is a cyclic group of order 2, then every permutation of the three subgroups of order 2 can be realized by an automorphism of G. If W is a cyclic group of order 2^i with $1 < i$, w an element of order 2^i in W, $z \neq 0$ an element of T, then all the automorphisms of G have the form $zf = z + z'$ with $n(z') \leq 1$, z' in W, $wf = cw + ez$ where c is an odd number and $e = 0$ or $= 1$, and therefore every automorphism of $L(G)$ is induced by an automorphism of G. If finally W is of type 2^∞, then every automorphism of G maps W upon itself and permutes the elements of order 2 in G which are not contained in W and therefore every automorphism of $L(G)$ is induced by an automorphism of G.

---

[13] The automorphisms of $L(G)$ preserve the situation of the subgroups of G (= situationstreue Abbildung von G); cf. R. Baer, "Situation der Untergruppen und Struktur der Gruppe," *Sitzungs-Berichte Heidelberg Akademie Wissenschaften, Math.-Nat. Kl.*, (2) (1933), pp. 12-17.

[14] The same property is possessed by, e.g., the subgroups of the additive group of the rational numbers.

[15] Cf. the paper, mentioned in footnote [13].

Assume now that $G = Z' + Z''$ where Z' and Z'' are cyclic groups, generated by elements z' and z'' respectively, neither of them of order 2. Then there exists an automorphism of $L(G)$ which

maps the subgroup generated by	upon the subgroup generated by
z'	$z' + p^{n(z'')-1}z''$
$z' + p^{n(z'')-1}z''$	$z' + p^{n(z'')-1}z'' + p^{n(z')-1}z'$
$z' + p^{n(z'')-1}z'' + p^{n(z')-1}z'$	z'

and all the other cyclic subgroups of G upon themselves. Such an automorphism of $L(G)$ cannot be induced by an automorphism of G.

A similar argument shows the existence of automorphisms of $L(G)$ which are not induced by automorphisms of $G = Z' + Z''$, if Z' and Z'' are both irreducible groups $\neq 0$ and not of order 2.

If finally $G = G' + G''$, then every automorphism of $L(G')$ is induced by an automorphism of $L(G)$ and therefore it follows from section 1 and the above considerations that the conditions of the theorem are necessary.

THEOREM 5.2. *The automorphism f of the primary Abelian group G of characteristic p induces the identical automorphism in $L(G)$ if, and only if, there exists an integral p-adic number $c \not\equiv 0 \bmod p$ such that $xf = cx$ for every element x in G.*[16]

It suffices to prove the necessity of the condition. If the automorphism f of G induces in $L(G)$ the identical automorphism, then f maps every cyclic subgroup of G upon itself and therefore

$$xf = c(x)x \text{ for every } x \text{ in } G.$$

Since f is an automorphism of G, the function $c(x)$ satisfies:

$$(c(x+y) - c(x))x = (c(y) - c(x+y))y.$$

If the cyclic groups, generated by x and y have no element $\neq 0$ in common, then

$$c(x+y) \equiv c(x) \bmod p^{n(x)}, \qquad c(x+y) \equiv c(y) \bmod p^{n(y)}$$

and consequently

$$c(x) \equiv c(y) \bmod p^{\text{minimum}(n(x),n(y))}.$$

[16] If $\sum_{i=0}^{\infty} c_i p^i$ is an integral p-adic number c, then $\sum_{i=0}^{j} c_i p^i x = \sum_{i=0}^{j+k} c_i p^i x$, if $n(x) \leq j \leq j + k$, and cx is therefore uniquely determined.

Since by section 1 $(G, n(x) \leq i)$ is a direct sum of cyclic groups, and since f induces an automorphism in $(G, n(x) \leq i)$, there exists by the above argument to every integer i an integer c_i such that

$$xf = c_i x \text{ for } n(x) \leq i$$

and this completes the proof, since

$$c_i \equiv c_{i-1} \bmod p^{i-1}.$$

6. The centralizer of the group of automorphisms.

Notations. $\boldsymbol{Z}(G)$ is the centralizer of the group $\boldsymbol{P}(G)$ of all the proper automorphisms of G, i. e. the group of all those proper automorphisms of G which are permutable with every proper automorphism of G.

$\boldsymbol{Z}^*(G)$ is the group of all those proper automorphisms f of G, satisfying:

$$Sf = S \text{ for every subgroup } S \text{ of } G.$$

It is a consequence of Theorem 5. 2 that

$$\boldsymbol{Z}^*(G) \leq \boldsymbol{Z}(G) \text{ for every primary Abelian group } G.$$

Theorem. *Assume that G is a primary Abelian group of characteristic p.*

(I) $\boldsymbol{Z}^*(G) \neq \boldsymbol{Z}(G)$ *if, and only if,*

(Ia) $p = 2$;

(Ib) *there exists a (uniquely determined) integer $m = m(G)$ such that*

$$1 = r(G, \infty) = r(G, m), \qquad 0 = r(G, k) \text{ for } m < k.$$

(II) *Assume that G satisfies* (Ia) *and* (Ib).

(IIa) *The proper automorphism f of G belongs to $\boldsymbol{Z}(G)$ if, and only if, $Sf = S$ for every subgroup S of G, satisfying*

$$S \leq (G, n(x) < m(G)) \text{ or } F(0) \leq S,$$

where $m(G)$ is the invariant of condition (Ib) *and $F(0) = (2^\infty G, n(x) \leq 1)$ is the group of all those elements of G which are fixed elements of every proper automorphism of G.*

(IIb) *$\boldsymbol{Z}(G)$ is the direct product of $\boldsymbol{Z}^*(G)$ and the group $\boldsymbol{W}$ of order 2 which is generated by the automorphism w of G, satisfying:*

$$xw \begin{Bmatrix} = x \\ = x + v \end{Bmatrix} \text{ if } x \begin{Bmatrix} \text{is} \\ \text{is not} \end{Bmatrix} \text{ an element of the join of } (G, n(x) < m(G))$$

and $2^\infty G$, where v is the element $\neq 0$ in $F(0)$.

Proof. 1. Assume that f belongs to $\boldsymbol{Z}(G)$.

If $G = D + G'$, then denote by g the automorphism of G, satisfying:

$$\begin{aligned} xg &= x \text{ for } x \text{ in } D, \\ &= -x \text{ for } x \text{ in } G'. \end{aligned}$$

If u is any element of D, then $uf = d + u'$ with d in D and u' in G' and

$$\begin{aligned} ugf &= uf = d + u' \\ &= ufg = dg + u'g = d - u', \end{aligned}$$

i. e. $2u' = 0$ and $2uf = 2d$ is an element of D and consequently:

(1) If f belongs to $\mathbf{Z}(G)$, D is a direct summand of G, then $2Df = 2D$.

2. Assume that f belongs to $\mathbf{Z}(G)$ and that $r(G, i) \neq 0$ for a certain integer i.

Then there exists by section 1 a decomposition

$$G = Z + G'$$

where Z is a cyclic group, generated by an element z of order p^i. If v is an element of G' such that $n(v) \leq i$, then $z + v$ generates a (cyclic) group Z' such that

$$G = Z' + G'.$$

Since, by (1), $2Zf = 2Z$, $2Z'f = 2Z'$, $2G'f = 2G'$, there exist integers c, c' such that

$$c'2z + c'2v = c'2(z + v) = 2(z + v)f = 2zf + 2vf = c2z + 2vf,$$

and since $2vf$ is an element of G', it follows that $c'2z = c2z$ and $2c \equiv 2c' \bmod p^i$ and therefore $2vf = 2cv$. Thus it has been proved:

(2) If f belongs to $\mathbf{Z}(G)$, then there exists for every integer i with $r(G, i) \neq 0$ an integer c_i such that

$$xf = c_i x \text{ for every } x \text{ in } 2(G, n(x) \leq i).$$

3. If f belongs to $\mathbf{Z}(G)$ and the orders of the elements of $G/p^\omega G$ are not bounded, then there exists an infinity of integers i such that $r(G, i) \neq 0$ and therefore every element x in $2G$ is mapped by f upon a multiple of x.

If the orders of the elements of $G/p^\omega G$ are bounded, then it follows from section 1, that

$$G = G' + p^\infty G,$$

where the orders of the elements in G' are bounded and G' is therefore a direct sum of cyclic groups. If p^m is the maximum order of the elements in G' and x an element in $p^\infty G$ such that $m < n(x)$, then x is contained in a direct summand S of type p^∞ of $p^\infty G$. Since $Sf = 2Sf = 2S = S$, there exists by (1) an integer $c(x)$ such that

$$xf = c(x)x \quad \text{and} \quad 2c(x) \equiv 2c_m \bmod p^m.$$

If g is any element of G, then $g = g' + g''$ with g' in G' and g'' in $p^\infty G$ and

$$\begin{aligned} 2g\boldsymbol{f} &= c_m 2g \quad \text{for} \quad n(g'') \leq n(g) \leq m, \\ &= c(g'')2g \quad \text{for} \quad m < n(g) = n(g'') \end{aligned}$$

and therefore every element x in $2G$ is mapped upon a multiple of x by $\boldsymbol{f}$. Thus it has been proved:

(3) If $\boldsymbol{f}$ belongs to $\boldsymbol{Z}(G)$, then $S\boldsymbol{f} = S$ for every subgroup S of $2G$.

Note that $2G = G$ for $p \neq 2$ and that therefore Theorem 5. 2 implies:

(3a) If $p \neq 2$, then the automorphism $\boldsymbol{f}$ of G belongs to $\boldsymbol{Z}(G)$ if, and only if, $S\boldsymbol{f} = S$ for every subgroup S of G.

4. Assume that $p = 2$ and that $\boldsymbol{f}$ belongs to $\boldsymbol{Z}(G)$. There exists by (3) and Theorem 5. 2, an integer 2-adic number $c = c(\boldsymbol{f})$ such that $2x\boldsymbol{f} = c2x$ for every element x in G.

Put $x\boldsymbol{h} = x\boldsymbol{f} - cx$ and this function $\boldsymbol{h}$ satisfies:

$n(x\boldsymbol{h}) \leq 1$ for every element x in G;
$(x + y)\boldsymbol{h} = x\boldsymbol{h} + y\boldsymbol{h}$, since $\boldsymbol{f}$ is an automorphism;
$x\boldsymbol{gh} = x\boldsymbol{gf} - cx\boldsymbol{g} = x\boldsymbol{fg} - cx\boldsymbol{g} = x\boldsymbol{hg}$ for every proper automorphism $\boldsymbol{g}$ of G, since $\boldsymbol{f}$ belongs to $\boldsymbol{Z}(G)$.

If $r(G, i) \neq 0$, then there exists by section 1 a decomposition:

$$G = Z + G'$$

where Z is a cyclic group, generated by the element z of order 2^i. If v is an element of G' such that $n(v) \leq i$, then a proper automorphism $\boldsymbol{g}$ of G is defined by

$$z\boldsymbol{g} = z + v, \qquad x\boldsymbol{g} = x \text{ for every } x \text{ in } G'.$$

Since $n(z\boldsymbol{h}) \leq 1$, $z\boldsymbol{h} = 2^s z + z'$ with $i - 1 \leq s \leq i$, z' in G' and therefore

$$\begin{aligned} z\boldsymbol{hg} &= 2^s z\boldsymbol{g} + z'\boldsymbol{g} = 2^s(z + v) + z' \\ &= z\boldsymbol{gh} = (z + v)\boldsymbol{h} = 2^s z + z' + v\boldsymbol{h}, \end{aligned}$$

and this proves:

(4) If $G = Z + G'$, where Z is a cyclic group of order 2^i, and $\boldsymbol{f}$ belongs to $\boldsymbol{Z}(G)$, then there exists an s such that $i - 1 \leq s \leq i$ and $x\boldsymbol{h} = 2^s x$ for every x in G' with $n(x) \leq i$.

If g is any element of G such that $n(g) < i$, then $g = 2kz + g'$ with g' in G' and $n(g') < i$, and consequently: $g\boldsymbol{h} = 0$ by (4). This proves:

(5) If $p = 2$, $r(G, i) \neq 0$, and $\boldsymbol{f}$ belongs to $\boldsymbol{Z}(G)$, then $x\boldsymbol{f} = cx$ for every x with $n(x) < i$.

If the orders of the elements of $G/2^\omega G$ are not bounded, then there exists

an infinity of integers i such that $r(G, i) \neq 0$ and therefore (5) and Theorem 5. 2 imply:

(6) If $p = 2$ and the orders of the elements of $G/2^{\omega}G$ are not bounded, then the proper automorphism $\boldsymbol{f}$ of G belongs to $\boldsymbol{Z}(G)$ if, and only if,

$$Sf = S \text{ for every subgroup } S \text{ of } G.$$

5. Assume now that $p = 2$, that the orders of the elements of $G/2^{\omega}G$ are bounded and that $\boldsymbol{f}$ belongs to $\boldsymbol{Z}(G)$.

Then there exists by section 1 a decomposition:

$$G = G' + 2^{\infty}G$$

where the orders of the elements in G' are bounded and G' is therefore a direct sum of cyclic groups. If 2^m is the maximum order in G', then in using the notation, introduced in 4, it follows from (5) that

$x\boldsymbol{f} = cx$ for every element x, contained in the join of $2^{\infty}G$ and $(G, n(x) < m)$.

$$G' = Z + T + G'',$$

where $G'' \leq (G, n(x) < m)$, T either 0 or a direct sum of cyclic groups of order 2^m and Z a cyclic group, generated by an element z of order 2^m. Then

$$z\boldsymbol{h} = 2^s z + z' \text{ with } z' \text{ in } T + G'' + 2^{\infty}G \text{ and } m - 1 \leq s \leq m$$

and (4) implies:

$$x\boldsymbol{h} = 2^s x \text{ for every element } x \text{ in } T + G'' + 2^{\infty}G.$$

If $T \neq 0$, then reasons of symmetry imply that $z' = 0$ and therefore:

(7) If $r(G, m) \neq 1$, $2^{\infty}G = 0$, then

$$x\boldsymbol{f} = (c + 2^s)x \text{ for every element } x \text{ in } G.$$

If $2^{\infty}G \neq 0$, then there exists in $2^{\infty}G$ an element w of order 2^m and $2^s w = w\boldsymbol{h} = 0$ implies that $s = m$ and therefore that $x\boldsymbol{f} = cx$ for every x in G and thus it follows from (7) and Theorem 5. 2:

(8) If $r(G, m) \neq 1$, then $\boldsymbol{f}$ belongs to $\boldsymbol{Z}(G)$ if, and only if, $S\boldsymbol{f} = S$ for every subgroup S of G.

If $r(G, m) = 1$, but $2^{\infty}G \neq 0$, then the above argument shows again that $s = m$, i. e. that

$$z\boldsymbol{h} = z' \text{ with } z' \text{ in } G'' + 2^{\infty}G.$$

If $\boldsymbol{g}$ is any automorphism of G, then

$$z\boldsymbol{g} = jz + z^* \text{ with } j \text{ an odd integer and } n(z^*) \leq m,\ z^* \text{ in } G'' + 2^{\infty}G.$$

Then

$$zgh = (jz + z^*)h = zh + z^*h = z', \text{ since } n(zh) \leq 1,$$
$$= zhg = z'g, \text{ i. e.}$$

zh is an element of $F(0)$. If conversely w is an element of $F(0)$ and the proper automorphism b is defined by

$$zb = cz + w, \qquad xb = cx \text{ for } x \text{ in } G'' + 2^\infty G,$$

where c is an integer 2-adic number $\not\equiv 0 \bmod 2$, then any proper automorphism g of G satisfies:

$$zbg = czg + wg = c(kz + z^*) + w \text{ with } k \text{ an odd number and } z^* \text{ in } (G'' + 2^\infty G, n(x) \leq m),$$
$$= k(cz + w) + cz^*, \text{ since } n(w) \leq 1 \text{ by Theorem 3.1,}$$
$$= kz\mathbf{b} + z^*b = zgb, \text{ and}$$
$$xbg = cxg = xgb \text{ for } x \text{ in } G'' + 2^\infty G, \text{ i. e.}$$

b belongs to $\mathbf{Z}(G)$. Thus it has been proved:

(9) If $r(G, m) = 1$, $2^\infty G \neq 0$, then the proper automorphism f of G belongs to $\mathbf{Z}(G)$ if, and only if, there exists an integral 2-adic number $c = c(f)$ such that

$$xf - c(f)x = 0 \text{ for every element } x \text{ in the join of } (G, n(x) < m) \text{ and } 2^\infty G,$$
$$\equiv 0 \bmod F(0) \text{ for every } x \text{ in } G.$$

If $r(G, m) = 1$, $2^\infty G = 0$, f an automorphism in $\mathbf{Z}(G)$, then, as proved above,

$$zh = 2^s z + z' \text{ with } z' \text{ in } G'', \text{ and } m - 1 \leq s \leq m.$$

If g is any automorphism of G, then $zg = jz + z^*$ with j odd and z^* in G'' and therefore

$$zgh = jzh + z^*h = 2^s z + z', \text{ since } n(z^*) < m,\ n(zh) \leq 1,$$
$$= zhg = 2^s zg + z'g = 2^s z + z'g, \text{ since } 2^s z \text{ is by Theorem 3.1 an element of } F(0), \text{ i. e.}$$

$z' = z'g$ is an element of $F(0)$. It follows now from Theorem 3.1 that $n(z') \leq 1$ and $z' = rz$ for a suitable r and thus it follows from Theorem 5.2 that

(10) If $r(G, m) = 1$, $2^\infty G = 0$, then the proper automorphism f of G belongs to $\mathbf{Z}(G)$ if, and only if,

$$Sf = S \text{ for every subgroup } S \text{ of } G.$$

6. Theorem 3.1, Theorem 5.2, (9) and (10) imply:

(11) If $r(G, m) = 1$, $r(G, \infty) \neq 1$, then the proper automorphism f of G belongs to $\boldsymbol{Z}(G)$ if, and only if,

$$Sf = S \text{ for every subgroup } S \text{ of } G.$$

7. Assume now that $1 = r(G, m) = r(G, \infty)$ and $p = 2$. Then it follows from Theorem 3.1 that

$$F(0) = (2^\infty G, n(x) \leq 1).$$

If f belongs to $\boldsymbol{Z}(G)$ and either $F(0) \leq S$ or $S \leq (G, n(x) < m)$, then it follows from (9) that

$$Sf = S.$$

If conversely the proper automorphism f satisfies:

$Sf = S$ for every subgroup S of G such that $F(0) \leq S$ or $S \leq (G, n(x) < m)$,

then there exists by Theorem 5.2 an integer 2-adic number c such that

$$xf = cx \text{ for every element } x \text{ in } 2G = \text{join of } (G, n(x) < m) \text{ and } 2^\infty G.$$

If furthermore $S = Z + F(0)$, where Z is a cyclic direct summand of G, generated by an element z of order 2^m, then $Sf = S$, i. e. $zf = jz + w$ where j is odd and w in $F(0)$. Since $2zf = c2z$, it follows that $2cz = 2jz$, i. e. $jz = (c + e2^{m-1})z$ with $e = 0, 1$. There exists furthermore in $2^\infty G$ an element u of order 2^{m+1}. Since w is an element of $F(0)$, since, by Theorem 3.1, $F(0) = (2^\infty G, n(x) \leq 1)$, and since $r(G, \infty) = 1$, it follows that $w = d2^m u$ where $d = 0$ or 1. The cyclic group, generated by $z + u$, contains $F(0)$ and it follows from the assumption, concerning the automorphism f, that

$$\begin{aligned}(z + u)f &= k(z + u) = kz + ku \text{ where } k \text{ is an odd integer,}\\ &= zf + uf = jz + w + cu\\ &= (c + e2^{m-1})z + (c + d2^m)u.\end{aligned}$$

This implies that

$$\begin{aligned}k &\equiv c + e2^{m-1} \bmod 2^m\\ &\equiv c + d2^m \bmod 2^{m+1} \text{ and therefore}\\ &\equiv c \bmod 2^m.\end{aligned}$$

Consequently $e = 0$ and $cz = jz$ and now it follows from (9) that f belongs to $\boldsymbol{Z}(G)$. This completes the proof of part (II) of the Theorem. Part (I) of the Theorem is a consequence of (3a), (6), (8) and (11).

Inspection of the automorphisms g, used in demonstrating the proposition (3) of the proof, shows that they are of order 2 and that therefore even the following statement holds true:

(12) If the automorphism f of the primary Abelian group G of characteristic p belongs to the kernel [17] $\boldsymbol{K}(G)$ of $\boldsymbol{P}(G)$, then $Sf = S$ for every subgroup S of $2G$. Thus, if $p \neq 2$, $\boldsymbol{K}(G) = \boldsymbol{Z}(G)$.

If $p = 2$, but the orders of the elements in G are not bounded, then there exist automorphisms in $\boldsymbol{Z}(G)$ of infinite order (cf. Theorem 5.2 and the fact that there exist integral 2-adic numbers whose multiplicative order is infinite) and it follows from a general theorem on kernels [18] that $\boldsymbol{K}(G) = \boldsymbol{Z}(G)$.

COROLLARY.[19] *The group $\boldsymbol{P}(G)$ of all the proper automorphisms of the primary Abelian group G is Abelian if, and only if, either G is irreducible or G is a direct sum of a group of order 2 and a group of type 2^∞.*

Remark. There exist Abelian groups without elements of finite order whose group of automorphisms is Abelian, e. g. the additive group of the rational numbers and its subgroups, the additive group of the integral p-adic numbers. An example of a finite non-commutative group whose group of automorphisms is Abelian has been given by Miller.

Proof. If G is irreducible, then every automorphism of G maps every subgroup of G upon itself and belongs therefore to the centralizer of $\boldsymbol{P}(G)$. If $G = T + G'$ where T is of order 2 and G' of type 2^∞, then every automorphism f of G satisfies: $Sf = S$ for every subgroup S of G such that $F(0) \leq S$, since $G'f = G'$ and $Tf \leq T + F(0)$. It follows therefore from the Theorem that $\boldsymbol{Z}(G) = \boldsymbol{P}(G)$, if G satisfies the condition of the Corollary.

If $G = Z + Y + G'$ where Z and Y are irreducible groups neither of them of order 2, then it follows from the Theorem that there exists an automorphism f of G which does not belong to $\boldsymbol{Z}(G)$ and this completes the proof.[20]

THE INSTITUTE FOR ADVANCED STUDY,
PRINCETON, N. J.

---

[17] The kernel of the (non-commutative) group Q consists of all the elements x in Q, satisfying $x^{-1}Sx = S$ for every subgroup S of Q, i. e. $x^{-1}yx = y^{h(y,x)}$ for every element y in Q. Cf. the paper, mentioned in footnote [3].

[18] Cf. Satz 3 of the paper, mentioned in footnote [2].

[19] This proposition has been proved for finite Abelian groups by A. Chatelet, "Les groupes abéliens finis et les modules des points entiers," *Travaux et mémoirs de l'Université de Lille*, nouv. sér. II, 3. Paris, 1925.

[20] In applying similar methods it is not difficult to prove the following proposition: *If G is a primary Abelian group and $Z(G) < P(G)$, then $P(G)/Z(G)$ is Abelian if, and only if,*

$$G = U + V$$

where U is of order 2 and V a cyclic group of order 2^i with $1 < i$.

POINTWISE PERIODIC HOMEOMORPHISMS.[1]

By Deane Montgomery.

1. In considering periodic transformations of locally Euclidean spaces there is suggested the question of whether or not there may exist a transformation under which each point is periodic but which is such that the periods of the points have no finite upper bound. The purpose of this paper is to show that such transformations do not exist, that is to show that any transformation under which each point is periodic is itself periodic. The known results on transformations of period p therefore apply for some p to any transformation under which each point is periodic.[2]

2. The space M considered here is a connected metric space[3] each of whose points is included in an open set homeomorphic to the interior of a solid n-dimensional sphere. If T is any transformation of M into itself T^k denotes the k-th iterate of T. It is understood that $T^0 = I$. A point x of M is periodic if there exists some positive integer k such that $T^k(x) = x$. The smallest positive integer k having this property is called the period of x. If each point of M is periodic under T, then T is said to be pointwise periodic. If there is some positive integer k such that $T^k = I$, the identity, then T is said to be periodic.

THEOREM. *If T is a pointwise periodic homeomorphism of M into itself then T is periodic.*

The proof is based on a lemma. It is convenient to make use of the integer valued function $p(x)$ defined to be the period of x under T. This function is defined everywhere in M and is lower semi-continuous. Let K be the set of points where $p(x)$ has an infinite least upper bound. The set K is closed.

---

[1] Received May 1, 1936.

[2] The principal contributions to the theory of periodic transformations have been made by Brouwer, Kerékjártó, Eilenberg, Newman, and P. A. Smith.

The author wishes to say that he has benefited from conversations with Leo Zippin on the topic of this note.

[3] Metric spaces were first used by Fréchet. See his *Les Espaces Abstraits.* See also Kuratowski, *Topologie I.*

LEMMA. *On every connected open set R in $M - K$, $p(x)$ is bounded.*

Since $p(x)$ is lower semi-continuous and since M is locally complete, $p(x)$ must be continuous on a set C which is everywhere dense in M;[4] C is an open set because $p(x)$ is integer valued. In the vicinity of a point of C $p(x)$ must be constant. There therefore exists a connected open set H in R such that $p(x)$ is constant and equal to an integer k everywhere in H. If H is not all of R let b be a point of R on the boundary of H. There is a neighborhood $U(b)$ in R and an integer N such that for any point x in $U(b)$, $p(x)$ is at most N. This is due to the fact that $p(x)$ has a finite least upper bound at b. Now consider the transformation $S = T^k$. Let

$$H_1 = \sum_{i=0}^{N!} S^i[H + U(b)].$$

This set is transformed homeomorphically into itself by S. Furthermore S is periodic of period at most $N!$, and for every x in H, $S(x) = x$. Since H_1 is connected it can be concluded from Newman's results[5] that every point in H_1 is fixed under S. But this means that if x is any point in H_1, then $T^k(x) = x$, so that every point of H_1 has period at most k. The set H_1 may, by the same process, be enlarged to a set H_2 (also in R) every point of which has period at most k. Continuing in this way by transfinite induction of any order necessary there is finally obtained a set H_α which coincides with R and every point of which has period at most k. This completes the proof of the Lemma.

3. The proof of the theorem will now be given by means of a *reductio ad absurdum*. Assume that $p(x)$ is unbounded. It follows from this assumption and from the lemma that $M - K$ is not connected and therefore that K is not vacuous. Let a be a point of K where $p(x|K)$[6] is continuous. There is an open set H including a such that $p(x|K)$ is constant and equal to some integer k everywhere in $H \cdot K$. Let R be any component of $M - K$ having points in H. There is by the lemma an integer N such that every point of R has period at most N. Now consider the set

$$W = \sum_{i=0}^{N!} S^i(R)$$

[4] Kuratowski, *loc. cit.*, p. 189.

[5] *Quarterly Journal of Mathematics*, vol. 2 (1931), p. 1.

[6] This notation is used to indicate the function whose domain of definition is K and which is equal to $p(x)$ everywhere on this set.

where $S = T^k$. If any point of $\bar{R} \cdot K \cdot H$ is an inner point of $W + K$, $p(x)$ is bounded at that point. Let c be any point of $\bar{R} \cdot K \cdot H$; by the remark just made c is not an inner point of $W + K$. Let V be a connected open set which is in H and which contains c. The set $V - \bar{W}$ is not vacuous. The set $V - W$ is separated from W by K and each point of $K \cdot H$ is fixed under S, so that the transformation S^* now to be defined is a homeomorphism of $V + W$ into itself. If x is in W, $S^*(x) = S(x)$ and if x is in $V - W$, $S^*(x) = x$. The set $V + W$ is connected and locally Euclidean so that Newman's results may again be applied, and it may be concluded that every point of $V + W$ is fixed under S^*. Therefore every point of W and in particular every point of R has period at most k. It follows, since R is an arbitrary component of $M - K$ having points in H, that every point of $(M - K) \cdot H$ has period at most k. But the same is true for every point of $H \cdot K$. Hence at points of $H \cdot K$ $p(x)$ has a finite least upper bound, and this contradicts the definition of K. The theorem is now demonstrated.

SMITH COLLEGE.

TRANSLATION GROUPS OF THREE-SPACE.[1]

By Deane Montgomery and Leo Zippin.

1. In this paper we seek conditions under which an abelian k-parameter group of transformations of an n-dimensional Euclidean space is essentially a k-parameter translation.[2] This problem is solved completely in case $k \geqq n-2$ or, in consequence, whenever $n = 3$. For $n > 3$ and general k, the problem is shown to be equivalent to an interesting question in topological product-spaces.

2. Let V^k denote the k-parameter (abelian) group of vectors of Euclidean k-dimensional space. The elements of V^k will be denoted by $u, v, w, \cdots$, the identity, in particular, by v_0. Let R denote an arbitrary metric space (which will subsequently be specialized). The points of R will be denoted by $x, y, z, \cdots$.

We shall say that V^k is a group of transformations of R provided that with each v there is uniquely associated a single-valued mapping of R into itself (*a priori*, not necessarily covering R) which we shall represent by

$$vx = x', \qquad x, x' \subset R;$$

furthermore these mappings must satisfy the following conditions:

i) $v_0 x = x$, for all x,

ii) $u(vx) = wx$, for all x, where $w = u + v$,

iii) vx is continuous in v and x simultaneously, that is if $v_n \to v$, and $x_n \to x$, then $v_n x_n \to vx$.[3]

If V denotes an arbitrary subset of V^k and X an arbitrary subset of R, then VX shall denote the set of all points in R of the form vx, where v is in V and x in X.

3. If x is any point of R, v any element of V^k, then for the element $y = (-v)x$ it follows from the first two conditions on V^k that $x = vy$. This implies that each v "covers" all of R and has, furthermore, a single-valued

[1] Received May 1, 1936.

[2] V. Niemytski has made such a study of one-parameter groups under the hypothesis of differentiability. See *Comptes Rendus de l'Académie des Sciences de Paris*, vol. 199 (1934), p. 18. Also, "Über Vollständige instabile dynamische Systeme," *Annali di Matematico* (4), vol. 14, pp. 275-286.

[3] Read "converges to."

inverse. The continuity condition assures that each V is a homeomorphism of R into itself.

Definition. A group of transformations V^k of R is called *dispersive* if with every pair of points x and y of R there is associated a pair of positive numbers ϵ and N such that for every point $z \subset S(x, \epsilon)$[4] and every v of norm greater than N, $|v| > N$, $vz \not\subset S(y, \epsilon)$.

4. THEOREM 1. *If V^k is a dispersive group of motions of R and R is locally compact, connected, and separable, then R may be represented as the topological product of two subspaces T_k and R_k:*

$$R = T_k \times R_k,$$

where T_k is homeomorphic to V^k.

We shall obtain the proof after a sequence of simple lemmas.

LEMMA 1. *If $x_n \to x$ and $v_n x_n \to y$, then $|v_n|$ is bounded.*

With each fixed $\epsilon > 0$ there is associated an integer $\bar{m}$ such that $x_n \subset S(x, \epsilon)$ and $v_n x_n \subset S(y, \epsilon)$, whenever $n \geqq \bar{m}$. For the ϵ and N which are associated with x and y by the dispersive character of V^k and for the corresponding $\bar{m}$, we have $|v_n| \leqq N$ whenever $n \geqq \bar{m}$.

LEMMA 2. *If $ux = vx$ for some point x, then $u = v$.*

Let $w = u - v$. Then $wx = (u - v)x = x$. Let $w_n = nw$, $n = 1, 2, \cdots$. Clearly $w_n x = x$. As a special case of Lemma 1, $|w_n|$ is bounded. Since $|w_n| = n|w|$, it follows that $|w| = 0$ and w must be the identity. That is, $u = v$.

LEMMA 3. *If $x_n \to x$ and $v_n x_n \to y$, then there is a v such that $v_n \to v$ and $y = vx$.*

Since v_n is bounded, by Lemma 1, there is necessarily at least one element v and a subsequence v_{i_n} such that $v_{i_n} \to v$. By the continuity, $v_{i_n} x \to vx$, so that $vx = y$. By Lemma 2, v is uniquely determined.

COROLLARY 1. *If V is any closed subset of V^k and x any point of R,* Vx is closed.

COROLLARY 2. *If $x_n \to x$ and $v_n x_n \to x$, then $v_n \to v_0$.*

LEMMA 4. *For fixed x, $V^k x$ is homeomorphic to the space V^k.*

The required correspondence between points of $V^k x$ and V^k is given by

[4] $S(x, \epsilon)$ denotes the set of points at a distance $\leqq \epsilon$ from x.

$$x' = v'x.$$

To each v', since x is fixed, there corresponds one and only one x'. To each x' of $V^k x$ there corresponds at least one v' such that $x' = v'x$ and, by Lemma 2, this v' is unique. The correspondence is therefore bi-univalued. That it is also bicontinuous is a consequence of Lemma 3 and the continuity iii).

4. Now let V denote a one-parameter closed subgroup of V^k. It follows, from definitions, that V is a dispersive group of transformations of R. Consequently all of the above Lemmas hold for the group V.

LEMMA 5. *Let v be any element of V, $v_t = tv$, $0 \leqq t \leqq c$, any constant. Then given any ϵ there exists a $\delta > 0$ such that if (x, y) [$=$ distance from x to y] $< \delta$, then $(v_t x, v_t y) < \epsilon$.*

This is a trivial consequence of the continuity of ux in u and x simultaneously.

Since Vx is homeomorphic to a straight line, by Lemma 4, it is a consequence of Lemma 5 that the space R is decomposed into a regular family of curves in the sense of H. Whitney. We shall use his work on cross-sections.[5] We define a *true-section*: a set of points Z of R is said to be a *true-section* (with respect to the regular family of curves) provided that

1) for every y of R the set $Z \cdot Vy$ consists of at most a single point,

2) if z_n is any sequence of points of Z and v_n any sequence of elements, $n = 1, 2, \cdots$, such that for some y of R, $v_n z_n \to y$ then there must exist a point z of Z and a subsequence z_{i_n} such that $z_{i_n} \to z$.

The second of these conditions implies, at once, that Z is closed. For if a point $z \subset \bar{Z}$, there is a sequence of points z_n of Z such that $z_n \to z$. Then $v_0 z_n = z_n \to z$, and 2) assures us that there is a $z' \subset Z$ and a subsequence z_{i_n} such that $z_{i_n} \to z'$. Clearly, $z' = z \subset Z$. Condition 2) implies also that VZ is closed. For if y is a limit point of VZ then for an appropriate choice of points z_n of Z and elements v_n, $v_n z_n \to y$. Certainly $v_{i_n} z_{i_n} \to y$, for any subsequence (i_n), $n = 1, 2, \cdots$. Then if $z_{i_n} \to z$, for some z of Z, we know from Lemma 3 that $y = vz$ for some v. It should be remarked that 2) is stronger than the requirement that Z and VZ be closed.[6] On the other hand, if Z is *compact* and closed it will obviously satisfy 2).

---

[5] H. Whitney, *Annals of Mathematics*, vol. 34 (1933), pp. 244-270. As Whitney points out in the last paragraph of his paper, cross sections may be obtained quite easily when, as here, the curves are given by a one-parameter group.

[6] The distinction can be made clear in a simple example. Let E be the Euclidean plane, V the group of translations parallel to the x-axis and Z the set of points consisting of the origin of coördinates together with the points of the hyperbola $xy = 1$. Here Z is *not* a true-section in the sense of condition 2.

A true section Z is said to be a true section at a point x provided that for all points y sufficiently near to x the intersection $Z \cdot Vy$ is not vacuous. It is said to "cover" a point y provided VZ covers y in the usual sense. It is clear that a true section at a point x "covers" x and also all points sufficiently close to x. Two true-sections Z_1 and Z_2 will be said to overlap provided the intersection $(VZ_1) \cdot (VZ_2)$ is not vacuous. A true-section Z "contains" a true-section Z' if VZ contains Z'. In case Z contains Z' in the customary sense, Z will be said to extend Z'.

5. Lemma 6. *True sections exist at every point x of R.*

Let W denote a Whitney cross-section at the point x.[7] In virtue of Corollary 2 to Lemma 3, there must exist a positive ϵ such that for all $y \subset S(x, \epsilon)$, W satisfies the first condition for a true-section. By reason of being a Whitney section W, for some ϵ' which may be supposed less than ϵ has the property that for all $y \subset S(x, \epsilon')$, $W \cdot Vy$ is not vacuous. Since R is supposed *locally compact*,[8] for some $\epsilon'' < \epsilon'$, $S(x, \epsilon'')$ is compact. Now, $Z = W \cdot S(x, \epsilon'')$ is a true-section at the point x.

Lemma 7. *Given two overlapping true-sections Z and Z', there exists a true section Z'' which "contains" Z and is an "extension" of Z'.*[9]

Since Z and VZ' are closed (see no. 4) it follows that $X = Z \cdot VZ'$ is closed. With each point x of X there is associated a unique element v and point x' of Z' such that $vx = x'$. The element v is a continuous function of x. For if x_n is a sequence of points of X such that $x_n \to x$, let $x'_n = v_n x_n$ be the corresponding points of Z'. Since $(-v_n)x'_n \to x$, it follows from condition 2) of no. 4 that a subsequence of the x'_n must converge to a point x', and this point must belong to Z' which is closed. By Lemma 3, $x' = vx$, where $v_n \to v$. Now v as a continuous *real-valued* function[10] of a closed subset X of a metric space Z can be extended to a continuous function over all of Z.[11] The set of points vz, where z ranges over Z and v is the associated functional value is easily seen to be the desired true section Z''.

6. We may proceed, finally, to a proof of Theorem 1. With each point x of R we may associate a true section Z_x. Each Z_x "covers" the corresponding point x. Since R is assumed separable, there will exist a countable sequence $Z_1, Z_2, \cdots$, of these sets such that they "cover" space: i. e. such that VZ_n,

---

[7] *Loc. cit.*, p. 256.

[8] This is the only point at which the local compactness of R is used.

[9] See definitions in no. 4.

[10] Since V is a one-parameter vector group.

[11] See for example Alexandroff and Hopf, *Topologie* I, p. 73.

$n = 1, 2, \cdots$, together cover R. Because R is connected we may assume that these sections are ordered in such a way that VZ_{n+1} has a non vacuous intersection with $\sum_{i=1}^{n} VZ_i$. By Lemma 7 there exists a true-section Z^1 which "extends" Z_1, and "contains" Z_2. By the same lemma there exists a true-section Z^2 which "extends" Z^1 and contains Z_2. Proceeding inductively we construct a set R_1 which is the point set sum of the monotonic increasing sequence of true-sections $Z^1, Z^2, Z^3, \cdots$. It is immediate that R_1 satisfies condition 1) of a true-section. Now suppose that y is any point such that for a sequence of points z_n of R_1 and a proper choice of elements v_n, $v_n z_n \rightarrow y$. Let Z_k be a true section in our first sequence which covers y. Then, for sufficiently large n, it will also cover all z_n. Now Z_k is "contained" in Z^k and Z^k must contain almost all z_n. But since Z^k is a true-section condition 2) must be satisfied for it and therefore condition 2) is also satisfied for R_1. It follows that R_1 is a true section.

It is trivial that R may be expressed as the topological product of V and R_1: $R = V \times R_1$ since each point of R is in the form vz for some unique element v and unique point z of R_1. Since R_1 is a closed subset of R, it is itself a locally compact separable metric space.

Now if $k = 1$, in our original theorem, the proof is complete. If $k > 1$, let V^{k-1} be a vector subgroup of V^k orthogonal to the original V of the argument above. We shall show that it is possible to interpret V^{k-1} as a dispersive group of motions over R_1. To this end, let v be an arbitrary element of V^{k-1}, x be any point of R_1 and $y = vx$ the corresponding point of R. The point y belongs to one and only one Vz where z is a point of R_1. We define a new correspondence (essentially a "projection" of the original one)

$$v: \quad \bar{v}x = z.$$

Here the point z is of the form ty for a unique element t of the group V.

That these new correspondences satisfy the conditions i) and iii) for a group of motions over R_1 (see no. 2) is clear. That they satisfy ii) follows from the obvious fact that $\bar{u}(\bar{v}x) \subset V\bar{w}x$, where the associated $w = u + v$. Since R_1 is a true section this implies that $\bar{u}(\bar{v}x) = \bar{w}x$, both being points of R_1. We have left to verify that V^{k-1}, thus interpreted, is dispersive. Let x and y denote arbitrary points of R_1 and let ϵ and N be chosen in accordance with the definition in no. 3. Now if there is any point z in $S(x, \epsilon)$ and any v of V^{k-1} such that the $\bar{v}z \subset S(y, \epsilon)$ then, for the appropriate t of V, $(t + v)z \subset S(y, \epsilon)$. But then, $| t + v | \leqq N$ and therefore $| v | \leqq N$ since v and t are orthogonal.

If we now choose in V^{k-1} a closed one-parameter vector group V', it follows by what has already been shown that

$$R_1 = V' \times R_2,$$

for some appropriately constructed "true-section" R_2 of R_1. By a finite induction, and because of the associativity of topological products, $R = V^k \times R_k$, for the corresponding R_k.

Since V^k is k-dimensional and homeomorphic to a subset of R it follows that $k \leqq \dim R$. In case $\dim R = k$, it follows from a theorem due to Hurewicz[12] that $\dim R_k = 0$.

7. *Definition.* We shall say that a V^k is a k-parameter translation group of a Euclidean space E_n if V^k is a group of motions of E_n and if there is a coördinate system $x_1, \cdots, x_n$ in E_n and a generating set of k vectors $v_1, v_2, \cdots, v_k$ in V^k such that to the vector rv_i, where r is an arbitrary real number, there corresponds the translation:

$$x'_j = x_j + \delta_j{}^i r x_j, \qquad (j = 1, 2, \cdots, n).$$

THEOREM 2. *The following three propositions are equivalent, n being arbitrary but fixed:*

A) A dispersive group of motions V^k of an E_n is a k-parameter translation group of E_n for all k;

B) If $E_n = E_k \times R_k$, then R_k is homeomorphic to E_{n-k}, for all k;

C) If $E_n = E_1 \times R_1$, then R_1 is homeomorphic to E_{n-1}.

Proof. That C) implies B) is a consequence of the associativity of topological products. That B) implies A) is a consequence of our Theorem 1. For from this theorem it follows that $E_n = V^k \times R_k$ and, since V^k is homeomorphic to E_k, B) implies that R_k is homeomorphic to E_{n-k}. Since this is the case we can introduce a coördinate system into E_n having the desired properties.

Finally, A) implies C). Suppose we have $E_n = E_1 \times R_1$. We may regard E_1 as a one-parameter vector group, some point e_0 corresponding to the identity vector and some point e_1 to the generating vector. Let z_0 be an arbitrary fixed point in R_1. Then E_1 and R_1 may be interpreted as subspaces of E_n, R_1 being the set of all points (e_0, z) where z ranges over R_1 and E_1 being the set of all points (e, z_0) where e ranges over E_1. Now corresponding to the vector e of E_1 we have in E_1 a certain translation; e: $e'(e) = e' + e$. We can extend this to a motion of all of E_n, the point (e', z) going over into the point $[(e' + e), z]$. It is quite easy to verify that E_1 now becomes a dispersive group of motions of E_n. Now, applying A) we get a coördinate

[12] W. Hurewicz, *Annals of Mathematics*, vol. 36 (1935), p. 194.

system in E_n such that the vectors of E_1 correspond to a change in the first coördinate only. Furthermore any desired change in the first coördinate can be accomplished by a suitable vector e of E_1. From this fact we may conclude that between the points of R_1 and the coördinates $(x_2, x_3, \cdots, x_n)$ there is a bi-univalued and bicontinuous correspondence.

Given any point (e_0, z) of R_1 we associate with it the set of its last $n-1$ coördinates: $x_2, \cdots, x_n$. No other point (e_0, z') of R_1 can have the same last coördinates for otherwise there would exist an element e of E_1 such that $[(e_0 + e), z']$ coincided with (e_0, z) (by suitably changing the first coördinate). But this contradicts the fact that $[(e_0 + e), z']$ cannot be a point of R_1 unless e is the identity e_0. On the other hand to any set of coördinates $(x_2, \cdots, x_n)$ there corresponds the unique point $(0, x_2, \cdots, x_n)$ and to this point (e, z) of E_n a unique point (e_0, z) of R_1. That the given correspondence is bi-univalued and bicontinuous should now be obvious.

Regarding C) we may remark that it is easy to establish a number of topological properties of R_1 which tend to corroborate the conjecture that it is indeed an E_{n-1}. Only in the case that $n \leqq 3$ are we able to prove this. We shall call attention here to the fact, which we shall have occasion to use in a moment, that if $E_n = E_k \times R_k$, then R_k is a locally compact separable metric space, connected and locally connected.

8. THEOREM 3. *A necessary and sufficient condition that a V^k $(k \geqq n-2)$ be a k-parameter translation group of E_n is that it be a dispersive group of motions of E_n.*

The necessity is quite obvious; we prove the sufficiency. Let x be any point of E_n and $V^k x$ the corresponding closed subset determined by the dispersive group V^k. This set is homeomorphic to an E_k. It is clear that if we adjoin to E_n the point at infinity, it becomes an n-sphere E^*_n and $V^k x$ becomes a k-sphere E^*_k. Then if $k = n-1$, E^*_k separates E^*_n into two connected sets and if $k = n-2$, E^*_k does not separate E^*_n (by the Alexander Duality Theorem). We can conclude at once that if $k = n-1$, $V^k x$ separates E_n and if $k = n-2$ it does not. Now we know that E_n can be expressed as $V_k \times R_k$. From what we have said it follows that for $k = n-1$, each point of R_k separates R_k into two connected sets and from this it follows that R_k is homeomorphic to an E_1.[13] In this case, compare no. 7, our theorem is proved.

Now if $k = n-2$, we see that no point of R_k is a cut-point and therefore R_k contains simple closed curves.[14] Let J be any simple closed curve of R_k.

---

[13] See, e.g., R. L. Moore, "Foundations of point-set theory," *American Mathematical Society Colloquium Publications*, vol. 13 (1932), p. 103 ff.

[14] R. L. Moore, *loc. cit.*, p. 124.

The set V^kJ is homeomorphic to the product of a circle by an E_k and it is easily seen that regarded as a subset of E^*_n it is the homeomorph of an absolute $(n-1)$-cycle. The cycle in question is obtained from the topologic product of a circle and an $(n-2)$-cube in which the point set corresponding to the product of the circle by the $(n-3)$-sphere bounding the cube has been "identified" with a single point. We conclude, by the Duality Theorem, that the set V^kJ separates E_n. Let x' and y' be two points in different components of $E_n - V^kJ$. If the points x and y of R_k belonging to V^kx' and V^ky' respectively belonged to the same component C of $R_k - J$, it would follow that x' and y' belonged to the connected set V^kC in $E_n \dot{-} V^kJ$. The contradiction shows that J separates R_k. That no arc L of J separates R_k is a consequence of the Duality Theorem and the fact that V^kL contains no absolute $(n-1)$-cycle. We may conclude that R_k is a "cylinder-tree": [15] i. e. that it is homeomorphic to the point set $S_2 - B$ where S_2 is the 2-sphere and B some closed and totally disconnected subset of S_2. Since R_k is obviously not compact, it is clear that B is not vacuous. All that remains to show is that B cannot contain more than a single point. Now if B contains as many as two points, it is clear that $S_2 - B$ contains some simple closed curve J' such that J' does not bound any singular 2-cell on $S_2 - B$. In this case the corresponding simple closed curve J of R_k cannot bound any singular 2-cell of R_k. But it is clear that if J is any simple closed curve of R_k it must certainly bound a singular 2-cell C' in E_n. Now each point of V^kC' determines a unique point of R_k. The set of these points is obviously a continuous image of C' and therefore itself a singular 2-cell which is a subset of R_k. The contradiction completes the demonstration that R_k is homeomorphic to an E_2. In view of no. 7, our theorem is proved for $k = n - 2$.

The case $k = n$ is instantly disposed of. In this case $E_n = V^n \times R_n$. If x denotes any point of R_n, V^nx being n-dimensional must contain inner points of E_n [16] and it follows at once that V^nx is open. Since it is also closed it must coincide with E_n. The set R_n is the single point x which obviously corresponds to the origin of the n-parameter translation group.

Finally we may state the obvious corollary to this theorem that for $n = 3$ any V^k (k necessarily $\leqq n$) of motions of E_n is a k-parameter translation group of E_n if and only if it is dispersive.

SMITH COLLEGE,
INSTITUTE FOR ADVANCED STUDY.

---

[15] L. Zippin, *American Journal of Mathematics*, vol. 52 (1930), pp. 331-350.
[16] K. Menger, *Dimensionstheorie*, Teubner, 1928.

DIFFERENTIAL PROPERTIES OF ABSTRACT TRANSFORMATION GROUPS WITH ABSTRACT PARAMETERS.[1]

By A. D. MICHAL and V. ELCONIN.

Introduction. Our main object in this paper is to present first-order differential equations which are satisfied by and, under specified conditions, characterize certain generalizations of abstract differentiable transformation groups. Our results clarify the classical theory of Lie, and abstract his first fundamental theorem and several related theorems to transformations in Banach spaces. A 'structural function,' analogous to the classical structural constants, appears in the condition of complete integrability for the fundamental differential equation and is shown to relate this equation to the differential equations of the so-called parameter groups in the same way that the structural constants relate the corresponding classical equations. Jacobi's identity is abstracted to Banach spaces and is used to establish an important property of the structural function. A differentiable group of matrix transformations and a group of linear functional transformations are used to illustrate the preceding results. Finally, several important unsolved problems are indicated, which arise when the differential equations are given and we try to determine when, in what domains, and with what group properties solutions of the equations exist.

The main definitions are collected in section 1. In section 2 the fundamental differential equation and the associated differential equations of the parameter groups are obtained, and conditions are given under which the solutions of these equations necessarily have group properties. The structural function is studied in section 3. The illustrative groups of transformations are in section 4.

1. *Definitions.* Let Δ, $\mathfrak{D}$ be any non-empty sets[2] and $=$ any reflexive, symmetric, transitive relation, such as logical identity, in Δ and in $\mathfrak{D}$.

DEF. 1. 1. A mark T is a *transformation of* $\mathfrak{D}$ if Tx is in $\mathfrak{D}$ for all x in $\mathfrak{D}$.

DEF. 1. 2. If T is a transformation of $\mathfrak{D}$, then T is *transitive over* $\mathfrak{D}$

[1] Presented to the Society, November 30, 1935 and April 11, 1936; received by the Editors, June 12, 1936.

[2] The sets Δ, $\mathfrak{D}$ will be in Banach spaces after Definition 1. 8.

if for any x in $\mathfrak{D}$, $x = Ty$ for some y in $\mathfrak{D}$; T is *simple over* $\mathfrak{D}$ if $Tx = Ty$ implies $x = y$ for any x, y in $\mathfrak{D}$; T is *simply transitive over* $\mathfrak{D}$ if it is simple and transitive over $\mathfrak{D}$.

DEF. 1. 3. If T, U are transformations of $\mathfrak{D}$, then $T = U$ *over* $\mathfrak{D}$ if

$$Tx = Ux \tag{1.1}$$

for all x in $\mathfrak{D}$; and TU, *the product of* T *and* U, is a transformation of $\mathfrak{D}$ satisfying

$$TUx = T(Ux) \tag{1.2}$$

for all x in $\mathfrak{D}$.

Transformation multiplication is evidently associative.

Let μ be in Δ. Let α' and $\alpha\beta$ be in Δ and T_α a transformation of $\mathfrak{D}$ for all α, β in Δ.

DEF. 1. 4. If

$$T_\beta T_\alpha = T_{\alpha\beta} \text{ over } \mathfrak{D} \tag{1.3}$$

for all α, β in Δ, then T_α is *closed over* Δ, $\mathfrak{D}$ *with respect to* $\alpha\beta$, and $\alpha\beta$ is a *parameter function over* Δ, $\mathfrak{D}$ *of* T_α.

DEF. 1. 5. If (1. 3) holds and

$$\alpha(\beta\gamma) = (\alpha\beta)\gamma \tag{1.4}$$

for all α, β, γ in Δ, then T_α *generates a semi-group over* Δ, $\mathfrak{D}$ *with respect to* $\alpha\beta$.

DEF. 1. 6. If

$$T_\mu x = x \tag{1.5}$$

for all x in $\mathfrak{D}$, then μ is *a unit in* Δ *over* $\mathfrak{D}$ *of* T_α.

DEF. 1. 7. If (1. 3), (1. 4), (1. 5) hold, then T_α *generates an integral semi-group over* Δ, $\mathfrak{D}$ *with respect to* $\alpha\beta$, μ.

DEF. 1. 8. If (1. 3), (1. 4), (1. 5) hold and

$$\alpha\mu = \alpha; \quad \alpha\alpha' = \mu; \quad \alpha\beta = \alpha\gamma \text{ implies } \beta = \gamma, \tag{1.6}$$

for all α, β, γ in Δ, then T_α *generates a group over* Δ, $\mathfrak{D}$ *with respect to* $\alpha\beta, \mu, \alpha'$.

(1. 4) and (1. 6) implies that Δ is a group, in the usual sense, with $\alpha\beta$ as the group operation, μ as unit element, and α' as the inverse of α.

Let T be the set of transformations T_α for α in Δ. If the conditions in

Definitions 1.5, 1.7, or 1.8 are satisfied, then T is respectively the semi-group, the integral semi-group, or the group which T_α generates, where in each case the product of elements in T is that defined in (1.2), T_μ is the unit element of T in the second and third cases, and $T_{\alpha'}$ is the inverse of T_α in the third case. Since we are more concerned here with analytic properties of $T_\alpha x$ than with set-theoretical properties of T, we keep T_α rather than T in evidence.

The theory of T is equivalent to a theory of a basic function $T(\alpha, x)$ on $\Delta, \mathfrak{D}$ to $\mathfrak{D}$, since there is a $1-1$ correspondence between the elements of T and the functions $T_\alpha x$ of x. However, we have preferred the transformation symbolism for several reasons, of which the most practical is the compact representation of iterations of the basic function furnished by transformation products.

Note that in the above definitions α, β can be replaced by any two distinct marks not already in use without altering the meaning of the defined propositions. Marks with this property will be called *umbrals.*

If the range and domains of $F(x, y, \cdots, z)$ are in Banach spaces[3] $d^{xy\cdots z}_{\xi\eta\cdots\zeta} F(x, y, \cdots, z)$ will denote the Fréchet differential[4] of $F(x, y, \cdots, z)$ with increments $\xi, \eta, \cdots, \zeta$. This form is occasionally ambiguous and we use $d^{xy\cdots z}_{\substack{\rho\sigma\cdots\tau\\ \xi\eta\cdots\zeta}} F(\rho, \sigma, \cdots, \tau)$ where $\rho, \sigma, \cdots, \tau$ are umbrals. $d^x_\xi F(x, y, \cdots, z)$ is sometimes more conveniently written as $F(x, y, \cdots, z; \xi)$. Successive differentiation with respect to the first argument x may accordingly be expressed by the adjunction of a succession of semi-colons and increments.

Now suppose Δ, $\mathfrak{D}$ are domains (open sets) in Banach spaces Σ, S.

DEF. 1.9. T_α is *continuous in* α *over* Δ, $\mathfrak{D}$ or *continuous over* Δ, $\mathfrak{D}$ according as $T_\alpha x$ is continuous in α or in α, x for all α; x in Δ, $\mathfrak{D}$.

T_α is *differentiable in* α over $\Delta, \mathfrak{D}$ or *differentiable over* Δ, $\mathfrak{D}$ *according* as $d^\alpha_\xi T_\alpha x$ exists or $d^{\alpha x}_{\xi z} T_\alpha x$ exists for all α, x in Δ, $\mathfrak{D}$.

DEF. 1.10. If T_α is differentiable in α over Δ, $\mathfrak{D}$, then α is *essential over* Δ, $\mathfrak{D}$ *in* T_α if for any function $\eta(\alpha)$ on Δ to Σ

$$d^{\alpha}_{\eta(\alpha)} T_\alpha x = 0$$

for all α, x in Δ, $\mathfrak{D}$ implies $\eta(\alpha) = 0$ for all α in Δ.

DEF. 1.11. A function F whose range and domains are in Banach

[3] S. Banach, *Théorie des Opérations Linéaires* (1932).

[4] M. Fréchet, *Annales Sc. Ec. Norm. Sup.*, t. 42 (1925), pp. 293-323. See also T. H. Hildebrandt and L. M. Graves, *Transactions of the American Mathematical Society*, vol. 29 (1927); A. D. Michal, *Annali di Matematica* (in press).

spaces is *definite* in one of its arguments with respect to the others if $F = 0$ for all values of the latter implies that the former $= 0$.

In Definitions 1.9, 1.10, α is an umbral.

2. *The fundamental differential equation.* Let $\Delta, \mathfrak{D}$ be domains in Banach spaces Σ, S. Let $\alpha\beta, \alpha'$ be in Δ and let T_α be a transformation of $\mathfrak{D}$ for all α, β in Δ. For brevity below, let α, β, γ be in Δ; ξ, η in Σ; x in $\mathfrak{D}$; z in S.

THEOREM 2.1. *If*

(i) *T_α is closed over $\Delta, \mathfrak{D}$ with respect to $\alpha\beta$,*

(ii) *T_α is differentiable over $\Delta, \mathfrak{D}$ and $d^x_z T_\alpha x$ is solvable in z for all α, x,*

(iii) *$\dfrac{\beta}{\alpha}$ exists in Δ, differentiable in α, and satisfying*

$$\alpha \frac{\beta}{\alpha} = \frac{\alpha\beta}{\alpha} = \beta$$

for all α, β, then for all α, x, ξ

$$d^\alpha_\xi T_\alpha x = U\big(T_\alpha x, V(T_\alpha x, \Omega(\alpha, \xi))\big), \tag{2.1}$$

where $U(x, z)$, $V(x, \xi)$, $\Omega(\alpha, \xi)$ are on $\mathfrak{D}, S$ to S; $\mathfrak{D}, \Sigma$ to S; and on Δ, Σ to Σ respectively, are linear in their second arguments, and $U(x, z)$ is solvable in z.

Proof. Let

$$d^{\alpha x}_{\xi z} T_\alpha x = G(\alpha, x, \xi) + H(\alpha, x, z).$$

Then

$$d^\alpha_\xi T_{\beta/\alpha} T_\alpha x = G\left(\frac{\beta}{\alpha}, T_\alpha x, d^\alpha_\xi \frac{\beta}{\alpha}\right) + H\left(\frac{\beta}{\alpha}, T_\alpha x, d^\alpha_\xi T_\alpha x\right) = 0,$$

since $T_{\beta/\alpha} T_\alpha = T_\beta$ over $\mathfrak{D}$. Let $P(\alpha, x, z)$ be the inverse with respect to z of $-H(\alpha, x, z)$. Then, with δ in Δ,

$$d^\alpha_\xi T_\alpha x = P\left(\frac{\beta}{\alpha}, T_\alpha x, G\left(\frac{\beta}{\alpha}, T_\alpha x, d^\alpha_\xi \frac{\beta}{\alpha}\right)\right) = P\left(\delta, T_\alpha x, G\left(\delta, T_\alpha x, d^\alpha_\xi{}^\sigma \frac{\alpha\delta}{\sigma}\right)\right),$$

where β has been replaced by $\alpha\delta$, since the middle member is obviously constant in β. Equation (2.1) now follows, with

$$U(x, z) = P(\delta, x, z), \quad V(x, \xi) = G(\delta, x, \xi), \quad \text{and} \quad \Omega(\alpha, \xi) = d^\alpha_\xi{}^\sigma \frac{\alpha\delta}{\sigma}.$$

The solvability in z of $U(x, z)$ is evident from that of $H(\alpha, x, z)$.

COROLLARY 2.1. *If in addition to the hypotheses of Theorem 2.1, T_α is transitive over $\mathfrak{D}$ and $d^\alpha_\xi T_\alpha x$ is solvable in ξ for all α, x, then $V(x, \xi)$, $\Omega(\alpha, \xi)$, and $V(x, \Omega(\alpha, \xi))$ are solvable in ξ.*

Proof. The solvability in ξ of $V(x,\xi)$ is evident. That of

$$U\big(x, V(x, \Omega(\alpha,\xi))\big)$$

follows from (2.1) and the transitivity of T_{α}. The remaining conclusions are consequences of Lemma 2.1 below, which can be stated briefly with the aid of

DEF. 2.1. Let E, F, G be Banach spaces. A mark L is *a linear transformation of E to F* if Lx is a linear function on E to F. If L, K are linear transformations of E to F and of F to G, then KL is the linear transformation of E to G satisfying

$$KLx = K(Lx)$$

for all x in E; if Lx is solvable and $L'x$ is its inverse, then L is *solvable* and the linear transformation L' on F to E is the *inverse of L.*

LEMMA 2.1. *If k, n are positive integers, $k \leq n$; E_i is a Banach space for $i = 1, 2, \cdots, n+1$; for $i = 1, 2, \cdots, n$, L_i is a linear transformation of E_i to E_{i+1} and L'_i, the inverse of L_i, exists if $i \neq k$; then L_k is solvable if and only if $L = L_n L_{n-1} \cdots L_1$ is solvable.*

Proof. If L_k is solvable, so is L, since the product of solvable linear transformations is solvable. If L is solvable, so is L_k, since

$$L'_k = L_{k-1}L_{k-2} \cdots L_1 L' L_n L_{n-1} \cdots L_{k+1}.$$

COROLLARY 2.2. *If* (i) *and* (ii) *of Theorem* 2.1 *are satisfied, and for all α, β*

$$\alpha(\alpha'\beta) = \alpha'(\alpha\beta) = \beta$$

and $\alpha'\beta$ is differentiable in α, then the conclusion of Theorem 2.1 *holds.*

Proof. In the proof of Theorem 2.1, replace $\dfrac{\beta}{\alpha}$ by $\alpha'\beta$.

Let $\alpha\beta = \Lambda_\beta\alpha = \Pi_\alpha\beta$. Then Λ_β and Π_α are transformations of Δ.

THEOREM 2.2. *If*

(i) *Δ is a group with $\alpha\beta$ as group operation, μ as unit element, and α' as the inverse of α for all α,*

(ii) *$\alpha\beta$ and α' are differentiable in α for all α, β, then $\alpha\beta$ is differentiable in β for all α, β; Λ_β and Π_α generate groups over Δ, Δ with respect to $\alpha\beta, \mu, \alpha'$; β and α are essential over Δ, Δ in Λ_β and Π_α respectively; Λ_β and Π_α are simply transitive over Δ for all α, β.*

Proof. Since $\alpha\beta = (\beta'\alpha')'$ and $\alpha\beta, \alpha'$ are differentiable in α, $\alpha\beta$ is dif-

ferentiable in β. Λ_β and Π_α are closed over Δ, Δ with respect to $\alpha\beta$; hence the second conclusion follows from the group properties of Δ, as does the fourth conclusion. The third conclusion follows from the solvability and differentiability of $\alpha\beta$ in α and in β, and from

LEMMA 2.2. *If a, b are in Banach spaces E, F; the functions $G(x)$, $H(y)$ are on a neighborhood of a to F and on a neighborhood of b to E; $d^a_\mu G(a)$ and $d^b_\nu H(b)$ exist; and*

$$H(G(x)) = x, \qquad G(H(y)) = y \tag{2.2}$$

for all x, y in neighborhoods of a, b; then the linear functions $d^a_\mu G(a)$, $d^b_\nu H(b)$ of μ, ν are mutually inverse.

Proof. Let $L\mu = d^a_\mu G(a)$, $K\nu = d^b_\nu H(b)$. Then differentiating equations (2.2) at $x = a$, $y = b$ gives

$$KL\mu = \mu, \qquad LK\nu = \nu$$

for all μ, ν in E, F.

Let T, Λ, Π be the sets of the transformations T_α, Λ_β, Π_α.

DEF. 2.2. If $\alpha\beta$ is a parameter function over $\Delta, \mathcal{D}$ of T_α, then Λ_β, Π_α are respectively the *first* and *second transformations* of Δ *associated with* T_α. If premise (i) of Theorem 2.2 holds, then Λ, Π are respectively the *first* and *second parameter groups associated with* T.

LEMMA 2.3. *If T_α is differentiable in α over $\Delta, \mathcal{D}$, α is essential over $\Delta, \mathcal{D}$ in T_α, the conclusion of Theorem 2.1 holds, and $\Omega(\alpha, \xi)$ is solvable in ξ for all α, then $U(x, V(x, \xi))$ is definite in ξ with respect to x.*

Proof. Suppose the conclusion is false. Then for some $\zeta \neq 0$ in Δ, $U(x, V(x, \zeta)) = 0$ for all x. Let $\eta(\alpha) = \Omega'(\alpha, \zeta)$, where $\Omega'(\alpha, \xi)$ is the inverse with respect to ξ of $\Omega(\alpha, \xi)$. Then $\eta(\alpha) \neq 0$, and from (2.1)

$$d^\alpha_{\eta(\alpha)} T_\alpha x = U\big(T_\alpha x, V(T_\alpha x, \Omega(\alpha, \eta(\alpha)))\big) = U(T_\alpha x, V(T_\alpha x, \zeta)) = 0$$

for all α, x, contrary to the essentiality of α in $T_\alpha x$.

THEOREM 2.3. *If*

(i) *T_α is closed over $\Delta, \mathcal{D}$ with respect to $\alpha\beta$, differentiable over $\Delta, \mathcal{D}$, transitive over $\mathcal{D}$ for all α, α is essential over $\Delta, \mathcal{D}$ in T_α, and $d^x_z T_\alpha x$ is solvable in z for all α, x,*

(ii) *α' is differentiable in α and such that for all α, β, ξ*

$$\alpha(\alpha'\beta) = \alpha'(\alpha\beta) = \beta, \qquad (\alpha\beta)' = \beta'\alpha', \qquad (\alpha')' = \alpha,$$

and $d_\xi^\alpha \alpha' \beta$ *exists and is solvable in* ξ, *then the conclusion of Theorem* 2.1 *holds, and for all* α, β, ξ, η, $\Omega(\alpha, \xi)$ *is solvable in* ξ, $\alpha\beta$ *is differentiable in* α *and in* β, *and*

$$d_\eta^\beta \alpha\beta = \Omega'(\alpha\beta, \Omega(\beta, \eta)), \tag{2.3}$$

$$d_\xi^\alpha \alpha\beta = \omega\big((\alpha\beta)', \Omega'\big((\alpha\beta)', \Omega(\alpha', \omega(\alpha, \xi))\big)\big), \tag{2.4}$$

where $\Omega'(\alpha, \xi)$ *is the inverse with respect to* ξ *of* $\Omega(\alpha, \xi)$, *and*

$$\omega(\alpha, \xi) = d_\xi^\alpha \alpha'.$$

Proof. By Corollary 2.2, the conclusion of Theorem 2.1 follows. From $(\alpha\beta)' = \beta'\alpha'$, $(\alpha')' = \alpha$, and the differentiability of α' and $\alpha'\beta$ in α follows the differentiability of $\alpha\beta$ in α and in β. The solvability of $\Omega(\alpha, \xi)$ in ξ follows from that of $d_\xi^\alpha \alpha'\beta$. Let $W(x, \xi) = U(x, V(x, \xi))$. Then from (2.1) and $T_\beta T_\alpha = T_{\alpha\beta}$ over $\mathfrak{D}$,

$$d_\eta^\beta (T_\beta T_\alpha x - T_{\alpha\beta} x) = W\big(T_{\alpha\beta} x, \Omega(\beta, \eta) - \Omega(\alpha\beta, d_\eta^\beta \alpha\beta)\big) = 0,$$

and from Lemma 2.3 and the transitivity of $T_{\alpha\beta}$ over $\mathfrak{D}$,

$$\Omega(\alpha\beta, d_\eta^\beta \alpha\beta) = \Omega(\beta, \eta),$$

from which (2.3) follows, and hence (2.4) on differentiating $\alpha\beta = (\beta'\alpha')'$.

THEOREM 2.4. *If*

(i) Λ_β *is closed over* Δ, Δ *with respect to* $\alpha\beta$, *differentiable over* Δ, Δ,

(ii) α' *is differentiable in* α *and such that for all* α, β

$$\alpha(\alpha'\beta) = \beta, \qquad (\alpha\beta)' = \beta'\alpha', \qquad (\alpha')' = \alpha,$$

then equations (2.3), (2.4) *hold and* β, α *are essential over* Δ, Δ *in* Λ_β, Π_α *respectively.*

Proof. From the first part of (i), $\alpha(\beta\gamma) = (\alpha\beta)\gamma$. Then from the second part of (ii), $\alpha'(\alpha\beta) = (\beta\alpha)\alpha' = (\beta\alpha')\alpha = \beta$. Hence $\alpha\beta$, $\alpha\beta'$ are mutually inverse in α : $(\alpha\beta')\beta = (\alpha\beta)\beta' = \alpha$. Similarly $\alpha\beta$, $\alpha'\beta$ are mutually inverse in β. By Lemma 2.2 and the second part of (i) $d_\xi^\alpha \alpha\beta$, $d_\eta^\beta \alpha\beta$ are solvable in ξ, η respectively, and hence β, α are essential over Δ, Δ in Λ_β, Π_α respectively. Moreover, from the first part of (ii) and the solvability of a product of solvable linear transformations, $d_\xi^\alpha \alpha'\beta = d^{\alpha'}_{\sigma \atop \omega(\alpha,\xi)} \sigma\beta$ is solvable in ξ. Hence (i) and (ii) of Theorem 2.3 are satisfied, and (2.3), (2.4) hold.

COROLLARY 2. 3. *If*

(i) Λ_β *is differentiable over* Δ, Δ *and generates a group over* Δ, Δ *with respect to* $\alpha\beta$, μ, α',

(ii) α' *is differentiable in* α, *then the conclusion of Theorem* 2. 4 *holds.*

Proof. (i) and (ii) of Theorem 2. 4 are evidently satisfied.

THEOREM 2. 5. *If*

(i) *for all* α, ξ, x : $W(x, \xi), \Omega(\alpha, \xi)$ *are in* $\mathfrak{D}, \Delta$ *respectively, are linear in* ξ, *and* $\Omega'(\alpha, \xi)$, *the inverse of* $\Omega(\alpha, \xi)$, *exists,*

(ii) *for all* $\alpha, \beta, \xi, \eta, x$

$$\begin{cases} d_\xi^\alpha T_a x = W(T_a x, \Omega(\alpha, \xi)) \\ T_\mu x = x, \end{cases} \tag{2. 5}$$

$$\begin{cases} d_\eta^\beta \Lambda_\beta \alpha = \Omega'(\Lambda_\beta \alpha, \Omega(\beta, \eta)) \\ \Lambda_\mu \alpha = \alpha, \end{cases} \tag{2. 6}$$

(iii) *the solution of* (2. 5) *is unique for each* x,

then T_a *is closed over* $\Delta, \mathfrak{D}$ *with respect to* $\alpha\beta$, μ *is a unit in* Δ *over* $\mathfrak{D}$ *of* T_a, *and if*

(iv) *the solution of* (2. 6) *is unique for each* α,

then T_a *generates an integral semi-group over* $\Delta, \mathfrak{D}$ *with respect to* $\alpha\beta$, μ.

Proof. Suppose (i), (ii), (iii). Then from (2. 5), (2. 6)

$$\begin{aligned} d_\eta^\beta T_{a\beta} x &= W\big(T_{a\beta} x, \Omega\big(\alpha\beta, \Omega'(\alpha\beta, \Omega(\beta, \eta))\big)\big) \\ &= W(T_{a\beta} x, \Omega(\beta, \eta)), \end{aligned}$$

since $\Omega'(\beta, \eta)$ is the inverse with respect to η of $\Omega(\beta, \eta)$. From (2. 5)

$$d_\eta^\beta T_\beta T_a x = W(T_\beta T_a x, \Omega(\beta, \eta)).$$

Moreover $T_{a\mu} = T_\mu T_a = T_a$ over $\mathfrak{D}$. Hence by (iii) the first conclusion follows. Now suppose (iv). From (2. 6)

$$\begin{aligned} d_\xi^\gamma \Lambda_{\beta\gamma} \alpha &= \Omega'\big(\Lambda_{\beta\gamma} \alpha, \Omega\big(\beta\gamma, \Omega'(\beta\gamma, \Omega(\gamma, \xi))\big)\big) \\ &= \Omega'(\Lambda_{\beta\gamma} \alpha, \Omega(\gamma, \xi)), \end{aligned}$$

and

$$d_\xi^\gamma \Lambda_\gamma \Lambda_\beta \alpha = \Omega'(\Lambda_\gamma \Lambda_\beta \alpha, \Omega(\gamma, \xi)).$$

Since also $\Lambda_{\beta\mu} = \Lambda_\mu \Lambda_\beta = \Lambda_\beta$ over Δ, (iv) implies $\Lambda_\gamma \Lambda_\beta = \Lambda_{\beta\gamma}$ over Δ, that is, $(\alpha\beta)\gamma = \alpha(\beta\gamma)$, and the last conclusion follows.

Theorems 2.1 and 2.5 together form an abstraction of Lie's first fundamental theorem.[5]

3. *The structural function* $\Gamma(\alpha, \xi, \eta)$. The definitions and notations of section 2 will apply throughout this section.

For all α, ξ, x, let $W(x, \xi)$, $\Omega(\alpha, \xi)$ be in $\mathfrak{D}, \Delta$ respectively, linear in ξ, and such that $d^x_z W(x, \xi)$, $d^\alpha_\eta \Omega(\alpha, \xi)$ exist, continuous in x and in α respectively. Then it can be shown[6] that the functions are totally differentiable in x, ξ and in α, ξ respectively. Suppose that $\Omega'(\alpha, \xi)$, the inverse with respect to ξ of $\Omega(\alpha, \xi)$, exists for all α, ξ, x. Then $\Omega'(\alpha, \xi)$ is linear[7] in ξ, differentiable[8] in α, and hence[6] totally differentiable in α, ξ.

Now if $d^x_z W(x, \xi)$ is continuous in x and T_a is transitive over $\mathfrak{D}$ and satisfies

$$(3.1) \qquad d^a_\xi T_a x = W(T_a x, \Omega(\alpha, \xi))$$

for all α, ξ, x, then, since $d^a_\xi d^a_\eta T_a x$ is symmetric[9] in ξ, η,

$$\nabla^\xi_\eta W\big(x, \Omega(\alpha, \xi);\ W(x, \Omega(\alpha, \eta))\big) = W(x,\ \nabla^\xi_\eta \Omega(\alpha, \eta; \xi))$$

or

$$(3.2) \qquad \nabla^\xi_\eta W(x, \eta; W(x, \xi)) = W(x, \Gamma(\alpha, \xi, \eta)),$$

where the effect of ∇^ξ_η is to subtract from its operand the result of interchanging ξ, η therein, and

$$(3.3) \qquad \Gamma(\alpha, \xi, \eta) = \nabla^\xi_\eta \Omega(\alpha, \Omega'(\alpha, \xi); \Omega'(\alpha, \eta))$$

is *the structural function over* Δ *of* T_a.

(3.2) is equivalent to the condition of complete integrability[10] for (3.1); this condition is therefore necessary for the existence of transitive solutions T_a of (3.1) when $W(x, \xi)$ is continuous in x, and it is also an important part of the least restrictive sufficient condition[11] known to us.

---

[5] S. Lie, *Transformationgruppen*, vol. I and vol. III.

[6] A. D. Michal, *Annali di Matematica, loc. cit.*; V. Elconin, Thesis (1937).

[7] J. Schauder, *Studia Mathematica*, vol. 2 (1930), pp. 1-6. See also S. Banach, *Studia Mathematica*, vol. 1 (1929), pp. 223-239.

[8] A. D. Michal and V. Elconin, "Completely integrable differential equations in abstract spaces." This memoir on existence theorems will appear in the *Acta Mathematica*.

[9] M. Kerner, *Annals of Mathematics*, vol. 34 (1933), pp. 546-572.

[10] A. D. Michal and V. Elconin, *Proceedings of the National Academy of Sciences*, vol. 21 (1935), pp. 534-536.

[11] A. D. Michal and V. Elconin, *Acta Mathematica, loc. cit.*

In the same way, if $d^{\alpha}_{\eta}\Omega(\alpha, \xi)$ is continuous in α and Λ_β is transitive over Δ and satisfies

$$d^{\beta}_{\eta}\Lambda_\beta\alpha = \Omega'(\Lambda_\beta\alpha, \Omega(\beta, \eta)) \tag{3.4}$$

for all α, β, η, then

$$\nabla^{\xi}_{\eta}\Omega'(\alpha, \eta; \Omega'(\alpha, \xi)) = \Omega'(\alpha, \Gamma(\beta, \xi, \eta)), \tag{3.5}$$

which is equivalent to the condition of complete integrability for (3.4); if moreover $(\alpha')' = \alpha$, $d^{\alpha}_{\xi} d^{\alpha}_{\eta}\alpha'$ exists, and

$$d^{\alpha}_{\xi}\alpha' = \omega(\alpha, \xi), \tag{3.6}$$

then, by differentiating $(\alpha')' = \alpha$: $\omega(\alpha, \xi)$, $\omega(\alpha', \xi)$ are mutual inverses with respect to ξ, and it is not difficult to show that if the conditions of complete integrability for (3.4) and (3.6) are satisfied so is that of

$$d^{\alpha}_{\xi}\alpha\beta = \omega\Big((\alpha\beta)', \Omega'\Big((\alpha\beta)', \Omega(\alpha', \omega(\alpha, \xi))\Big)\Big). \tag{3.7}$$

Several equations equivalent to (3.3) are sometimes useful; for example

$$\nabla^{\xi}_{\eta} d^{\alpha}_{\eta}\Omega(\alpha, \xi) = \Gamma(\alpha, \Omega(\alpha, \xi), \Omega(\alpha, \eta)), \tag{3.8}$$

an abstraction of Maurer's equations. Or, differentiating $\Omega(\alpha, \Omega'(\alpha, \xi)) = \xi$,

$$\Omega\Big(\alpha, \nabla^{\xi}_{\eta}\Omega'(\alpha, \eta; \Omega'(\alpha, \xi))\Big) = \nabla^{\xi}_{\eta}\Omega(\alpha, \Omega'(\alpha, \xi); \Omega'(\alpha, \eta)),$$

and hence

$$\Gamma(\alpha, \xi, \eta) = \Omega\Big(\alpha, \nabla^{\xi}_{\eta}\Omega'(\alpha, \eta; \Omega'(\alpha, \xi))\Big), \tag{3.9}$$

or

$$\nabla^{\xi}_{\eta}\Omega'(\alpha, \eta; \Omega'(\alpha, \xi)) = \Omega'(\alpha, \Gamma(\alpha, \xi, \eta)). \tag{3.10}$$

THEOREM 3.1.

(i) *The structural function* $\Gamma(\alpha, \xi, \eta)$ *is bilinear and alternating in* ξ, η.

(ii) *If* $W(x, \xi)$ *is definite in* ξ *with respect to* x *and the condition of complete integrability for* (3.1) *is satisfied, then* $\Gamma(\alpha, \xi, \eta)$ *is constant in* α.

(iii) *The condition of complete integrability for* (3.4) *is satisfied if and only if* $\Gamma(\alpha, \xi, \eta)$ *is constant in* α.

Proof. From (3.3), $\Gamma(\alpha, \xi, \eta)$ is evidently alternating in ξ, η, and additive in ξ and in η. The continuity in ξ and in η, and hence (i) follows from a theorem [12] of Banach on the limit of a sequence of additive functions.

[12] S. Banach, *Fundamenta Mathematicae*, vol. 3 (1922), pp. 133-181.

If (3.2) holds, $W(x, \Gamma(\alpha, \xi, \eta))$ is evidently constant in α, hence so is $\Gamma(\alpha, \xi, \eta)$ if $W(x, \xi)$ is definite in ξ with respect to x.

Since $\Omega(\alpha, \xi)$ is uniquely solvable in ξ, $\Omega(\alpha, \xi) = 0$ implies $\xi = 0$. Hence (iii) follows on comparing (3.5) and (3.10).

Another important property of $\Gamma(\alpha, \xi, \eta)$ can be established by means of an abstraction of Jacobi's Identity.

Let y be in S, θ in Σ, and for any function $F(x)$ on $\mathfrak{D}$ to S such that $d_y^x d_z^x F(x)$ exists, continuous in x, let

$$X_\xi F(x) = F(x; A(x, \xi))$$

where $A(x, \xi)$ is on $\mathfrak{D}, \Sigma$ to S, linear in ξ, and $d_y^x d_z^x A(x, \xi)$ exists, continuous in x.

Let

$$\begin{aligned}(X_\xi, X_\eta)F(x) &= \nabla_\eta^\xi X_\xi(X_\eta F(x)) \\ &= \nabla_\eta^\xi F(x; A(x, \xi); A(x, \eta)) + F\big(x; \nabla_\eta^\xi A(x, \eta; A(x, \xi))\big) \\ &= F\big(x; \nabla_\eta^\xi A(x, \eta; A(x, \xi))\big)\end{aligned}$$

since $d_y^x d_z^x F(x)$ is symmetric in y, z. Then Jacobi's Identity becomes

$$\Sigma((X_\xi, X_\eta), X_\theta)F(x) = 0, \tag{3.11}$$

where the summation is over the cyclical permutations of ξ, η, θ; for $((X_\xi, X_\eta), X_\theta)F(x)$ equals

$$\begin{aligned}&F\big(x; A(x, \theta); \nabla_\eta^\xi A(x, \eta; A(x, \xi))\big) \\ &-F\big(x; \nabla_\eta^\xi A(x, \eta; A(x, \xi)); A(x, \theta)\big) \\ &+F\big(x; \nabla_\eta^\xi A\big(x, \theta; A(x, \eta; A(x, \xi))\big)\big) \\ &-F\big(x; \nabla_\eta^\xi A\big(x, \eta; A(x, \xi; A(x, \theta))\big)\big) \\ &-F\big(x; \nabla_\eta^\xi A\big(x, \eta; A(x, \xi); A(x, \theta)\big)\big),\end{aligned}$$

in which the first and second terms cancel, the sums in (3.11) arising from the third and fourth terms cancel, and the sum arising from the fifth term equals zero.

THEOREM 3.2. *If*

(i) *$W(x, \xi)$ is definite in ξ with respect to x and*

(ii) *for all x, y, z, ξ $d_y^x d_z^x W(x, \xi)$ exists, continuous in x, or*

(iii) *for all $\alpha, \eta, \theta, \xi$ $d_\eta^\alpha d_\theta^\alpha \Omega'(\alpha, \xi)$ exists, continuous in α, then*

$$\sum_{\xi,\eta,\theta} \Gamma(\alpha, \Gamma(\alpha, \xi, \eta), \theta) = 0, \tag{3.12}$$

where the summation is cyclical.

Proof. Suppose (i), (ii), and $A(x,\xi)=W(x,\xi)$ for all x,ξ. Then by (3.2)

$$(X_\xi, X_\eta)F(x)=F\big(x;\nabla^\xi_\eta W(x,\eta;W(x,\xi))\big)$$
$$=F\big(x;W(x,\Gamma(\alpha,\xi,\eta))\big)=X_{\Gamma(\alpha,\xi,\eta)}F(x),$$

and hence

$$\sum_{\xi,\eta,\theta}((X_\xi,X_\eta),X_\theta)F(x)=X_JF(x)=0$$

by Jacobi's Identity, where J is the left member of (3.12). If in particular $F(x)=x$, for all x, then $X_JF(x)=W(x,J)=0$, and from (i), $J=0$.

Now suppose (iii), S is Σ, $\mathfrak{D}$ is Δ, and $A(\alpha,\xi)=\Omega'(\alpha,\xi)$ for all α,ξ. Then by (3.10) and Jacobi's Identity, with $F(x)=x$: $\Omega'(\alpha,J)=0$. Hence $J=0$, since $\Omega'(\alpha,\xi)$ is uniquely solvable in ξ.

4. *Illustrative groups.* The results in the preceding sections can be illustrated by a group of matrix transformations.

Let S,Σ be respectively the set of all matrices with m rows, n columns, and the set of all matrices with n rows, n columns, where in each set the matrix elements are in the field of all complex numbers. For any $\omega=(\omega_{ij})$ in S or in Σ, let $\|\omega\|=\max|\omega_{ij}|$. When $\|\omega\|$ as the norm of ω, and the usual definitions of matrix addition and multiplication, S,Σ are Banach spaces, and $x\alpha$, $\alpha\beta$ are bilinear functions on S,Σ to S and on Σ,Σ to Σ respectively.

Let $\mathfrak{D}$ be S, Δ the set of non singular matrices in Σ, μ the identity matrix in Σ, α' the inverse matrix of α for all α in Δ, and let $T_\alpha x=x\alpha$ for all α, x in $\Delta,\mathfrak{D}$. Clearly Δ is a domain. For brevity, let α,β be in Δ; ξ in Σ; x, z in $\mathfrak{D}$.

T_α generates a group over $\Delta,\mathfrak{D}$ with respect to $\alpha\beta$, μ, α'.

T_α is differentiable over $\Delta,\mathfrak{D}$, since for all α, x, ξ, z

$$d^{\alpha x}_{\xi z}T_\alpha x=x\xi+z\alpha.$$

$d^\alpha_\xi\alpha'$ exists equal to $-\alpha'\xi\alpha'$. This follows from a known theorem,[13] on noting that $\alpha\xi$, $\alpha'\xi$ are mutually inverse with respect to ξ. Or it may be deduced directly as follows. $(\alpha+\xi)'-\alpha'=-(\alpha+\xi)'\xi\alpha'$ if $\|\xi\|$ is sufficiently small. From the definition of matrix multiplication, $\|\alpha\beta\|\leq n\|\alpha\|\ \|\beta\|$, and, say from Kramer's Rule, $\|(\alpha+\xi)'\|$ is bounded for all sufficiently small $\|\xi\|$. Hence

$$\|(\alpha+\xi)'-\alpha'\|\leq n^2\|(\alpha+\xi)'\|\ \|\xi\|\ \|\alpha'\|$$

[13] A. D. Michal and V. Elconin, *Acta Mathematica, loc. cit.* This theorem asserts the differentiability with respect to a parameter of the inverse of a solvable linear function itself differentiable in the parameter.

is arbitrarily small for all sufficiently small $\| \xi \|$: α' is continuous in α; and for any positive number ϵ

$$\begin{aligned} \| (\alpha + \xi)' - \alpha' + \alpha'\xi\alpha' \| &= \| \{ (\alpha + \xi)' - \alpha' \} \xi\alpha' \| \\ &\leq n^2 \| (\alpha + \xi)' - \alpha' \| \; \| \alpha' \| \; \| \xi \| \leq \epsilon \| \xi \| \end{aligned}$$

if $\| \xi \|$ is sufficiently small. Since $-\alpha'\xi\alpha'$ is also linear in ξ, $d_\xi^\alpha \alpha' = -\alpha'\xi\alpha'$.

With the notations introduced in the proof of Theorem 2. 1,

$$U(x, z) = -z\delta', \qquad V(x, \xi) = x\xi, \qquad \Omega(\alpha, \xi) = -\alpha'\xi\delta,$$

since $\beta/\alpha = \alpha'\beta$ here. Hence by that theorem

$$d_\xi^\alpha T_a x = -(T_a x)(-\alpha'\xi\delta)\delta' = (T_a x)\alpha'\xi, \tag{4. 1}$$

which may be directly verified.

For simplicity, let $\delta = -\mu$. With the notations of Theorem 2. 3,

$$\Omega'(\alpha, \xi) = \alpha\xi, \qquad \omega(\alpha, \xi) = -\alpha'\xi\alpha'.$$

Hence by that theorem

$$d_\eta^\beta \alpha\beta = (\alpha\beta)(\beta'\eta) = \alpha\eta, \tag{4. 2}$$

and

$$d_\xi^\alpha \alpha\beta = -(\alpha\beta)\{(\alpha\beta)'\alpha(-\alpha'\xi\alpha')\}(\alpha\beta) = \xi\beta, \tag{4. 3}$$

which are evidently true.

With the notations of section 3,

$$\Gamma(\alpha, \xi, \eta) = \alpha' \nabla_\eta^\xi \alpha\eta\xi = \eta\xi - \xi\eta,$$

which is constant in α, and hence the condition of complete integrability for (4. 2) is satisfied. This is already evident since the transformation Λ_β defined by $\Lambda_\beta\alpha = \alpha\beta$ is transitive over Δ. For similar reasons, (4. 1) and (4. 3) are completely integrable. Moreover, the solutions of these equations are uniquely determined by one-point initial conditions. Thus T_a satisfies

$$\left\{ \begin{aligned} d_\xi^\alpha T_a x &= (T_a x)\alpha'\xi \\ T_\mu x &= x. \end{aligned} \right. \tag{4. 4}$$

Conversely, let $T_a x$ now be any solution of (4. 4). Then

$$d_\xi^\alpha (T_a x)\alpha' = (d_\xi^\alpha T_a x)\alpha' - (T_a x)(\alpha'\xi\alpha') = ((T_a x)\alpha'\xi)\alpha' - (T_a x)(\alpha'\xi\alpha') = 0.$$

Hence for some y in $\mathfrak{D}$, $(T_a x)\alpha' = y$, or $T_a x = ya$ for all α, and from $T_\mu x = x$ follows $x = y$. Essentially the same device may be used on (4. 2), (4. 3).

The above results all hold when the matrix transformation group is replaced by a certain group of linear functional transformations of the third kind.[14] Let S be the well-known Banach space of functions x^σ, real and continuous over $a \leqq \sigma \leqq b$, with $\| x^\sigma \| = \max_{a \leqq \sigma \leqq b} | x^\sigma |$. Let Σ be the set of all ordered pairs $(\alpha^\sigma, \alpha^\sigma_\tau)$ of functions α^σ and α^σ_τ, real and continuous for $a \leq \sigma$, $\tau \leq b$. Let

$$\| (\alpha^\sigma, \alpha^\sigma_\tau) \| = \max_{a \leq \sigma \leq b} | \alpha^\sigma | + \max_{a \leq \sigma,\, \tau \leq b} | \alpha^\sigma_\tau |$$

$$(\alpha^\sigma, \alpha^\sigma_\tau) + (\beta^\sigma, \beta^\sigma_\tau) = (\alpha^\sigma + \beta^\sigma, \alpha^\sigma_\tau + \beta^\sigma_\tau),$$

and

$$r(\alpha^\sigma, \alpha^\sigma_\tau) = (r\alpha^\sigma, r\alpha^\sigma_\tau),$$

for all $(\alpha^\sigma, \alpha^\sigma_\tau)$, $(\beta^\sigma, \beta^\sigma_\tau)$ in Σ and all real numbers r. Then Σ is a Banach space.[15] Let x, α, β be x^σ, $(\alpha^\sigma, \alpha^\sigma_\tau)$, $(\beta^\sigma, \beta^\sigma_\tau)$ respectively, and define

$$x\alpha = x^\sigma \alpha^\sigma + x^\rho \alpha^\sigma_\rho;$$
$$\alpha\beta = (\alpha^\sigma \beta^\sigma, \alpha^\sigma_\tau \beta^\sigma + \alpha^{(\tau)} \beta^\sigma_\tau + \alpha^\rho_\tau \beta^\sigma_\rho),$$

where in any term an index occurring once as subscript, once as superscript, and not in parentheses, indicates Riemann integration over (a, b) with respect to that index; thus $\alpha^\rho_\tau \beta^\sigma_\rho = \int_a^b \alpha^\rho_\tau \beta^\sigma_\rho d\rho$, but τ is free in $\alpha^{(\tau)} \beta^\sigma_\tau$. $x\alpha$, $\alpha\beta$ are bilinear functions on S, Σ to S and on Σ, Σ to Σ respectively. Let $T_a x = x\alpha$. Then T_a, a linear functional transformation of the third kind, generates a semi-group over Σ, S with respect to $\alpha\beta$. Let $\mathfrak{D}$ be S, Δ the set of all α in Σ for which $\alpha^\tau \neq 0$ and the Fredholm determinant $D[\alpha^\sigma_\tau / \alpha^{(\tau)}] \neq 0$, Γ^σ_τ the resolvent kernel of $\alpha^\sigma_\tau / \alpha^{(\tau)}$, and let $\mu^\sigma = 1$, $\mu^\sigma_\tau = 0$ for $a \leq \sigma$, $\tau \leq b$. Then[16] T_a generates a group over Δ, $\mathfrak{D}$ with respect to $\alpha\beta$, μ, α', where $\mu = (\mu^\sigma, \mu^\sigma_\tau)$, $\alpha' = (1/\alpha^\sigma, \Gamma^\sigma_\tau / \alpha^{(\sigma)})$. The results obtained for the matrix transformation group, and the arguments leading to them, except that employed in the direct deduction of $d^a_\xi \alpha' = -\alpha' \xi \alpha'$, all hold with the new interpretations of $T_a x$, $\alpha\beta$.

5. *Conclusion.* The main results of this paper express differential properties of abstract differentiable transformation groups and certain of their generalizations. The study of the intimate group structure, which we have

[14] A. D. Michal and T. S. Peterson, *Annals of Mathematics*, vol. 32 (1931), pp. 431-450. See also A. D. Michal, *American Journal of Mathematics*, vol. 50 (1928), pp. 473-517.

[15] A. D. Michal and R. S. Martin, *Journal de Mathématiques Pures et Appl.*, vol. 13 (1934), pp. 69-91.

[16] A. D. Michal and T. S. Peterson, *loc. cit.*

already begun, suggests that the totality of non-trivial differential properties is very large. It is at least reasonable to expect that salient formal group properties strongly imply and may sometimes be characterized by differential properties. The determinations of analytic conditions under which the latter properties imply the former constitute the most difficult and laborious problems in the theory. For example, in Theorem 2. 5, it was shown that the uniqueness of the solutions of certain differential equations implied fundamental formal group properties of these solutions, but the determination of reasonably broad analytic conditions which insure this uniqueness is difficult; our best results so far seem too restrictive. It is probable, however, that better conditions will be obtained by strengthening and extending basic existence theorems for completely integrable abstract differential equations which we have already established.[17]

CALIFORNIA INSTITUTE OF TECHNOLOGY,
PASADENA, CALIFORNIA.

[17] A. D. Michal and V. Elconin, *Acta Mathematica, loc. cit.*

REMARKS ON SYSTEMS OF ORDINARY DIFFERENTIAL EQUATIONS.[1]

By E. R. van Kampen.

I. *A uniqueness theorem.* Let the vector $\phi = \phi(\xi) = (f_1, \cdots, f_n)$ be a continuous function of the vector $\xi = (x_1, \cdots, x_n)$ in a certain region R of the ξ-space. Let t (or s) be a real variable and let a subscript t (or s) denote total differentiation with respect to this variable. It is well known that if ξ_0 is a fixed vector in R, then the system of ordinary differential equations

$$(1) \qquad \xi_t = \phi(\xi)$$

has at least one solution

$$(2) \qquad \xi = \psi(\xi_0, t), \text{ for which } \psi(\xi_0, 0) = \xi_0$$

and which is defined for all t in a certain interval which contains $t = 0$ and depends on ξ_0. Several conditions concerning the function $\phi(\xi)$ are known, each of which implies that the solution (2) of (1) is uniquely determined by ξ_0. In the uniqueness theorem[2] stated below, the same conclusion is drawn from an assumption on a certain set of solutions of (1), without any further restriction on the function $\phi(\xi)$ itself. Using the symbol $|\eta|$ to denote the length of the vector η, this uniqueness theorem may be formulated as follows:

Let the equation (1) *have a solution* (2) *such that*

(i) *one and only one curve of the system* (2) *passes through any point of* R, i. e.,

$$(3) \qquad \psi(\psi(\xi_0, t), s) = \psi(\xi_0, t + s),$$

whenever the expression on the left side is defined and

(ii) *the function* (2) *satisfies a Lipschitz condition with respect to* ξ_0, i. e., *there exsts a positive number* C *which is independent of* ξ_0 *and* η_0 *such that*

[1] Received November 20, 1936.

[2] This theorem combined with a treatment of Lie groups, obtained by the author several years ago but never published, implies that a group nucleus homeomorphic with an n-cell is a Lie group nucleus if (i) the functions $c^\gamma = f^\gamma(a^\alpha, b^\beta)$ defining the group operation have a continuous derivative with respect to a^α at a fixed value of a^α and (ii) the functions $f^\gamma(a^\alpha, b^\beta)$ satisfy a Lipschitz condition with respect to b^β. A publication of the considerations involved is planned in the near future.

(4) $$|\psi(\xi_0, t) - \psi(\eta_0, t)| < C\,|\xi_0 - \eta_0|$$

for any ξ_0 *and* η_0 *in* R.

Then, if ξ_0 *is a fixed vector in* R, (2) *defines the only solution of* (1) *which reduces for* $t = 0$ *to* ξ_0.

Let $\chi(t)$, $t_1 < t < t_2$, be any fixed solution of (1) and put

(5) $$\zeta(s, t) = \psi(\chi(s), t - s),$$

where $t_1 < s < t_2$, while, for a given s, the variable t is restricted in such a way that the right side of (5) has a meaning. Obviously (5) determines, for a fixed s, the unique curve of the system (2) through the point $\chi(s)$, so that the theorem is equivalent to the statement that $\zeta(s, t)$ is independent of s, whenever $\chi(s)$ is a solution of (1).

First, it is seen from (2) and (5) that $\zeta(t, t) = \chi(t)$, so that

(6) $$\zeta_t(t, t) = \phi(\zeta(t, t)),$$

since $\chi(t)$ is a solution of (1). Also, since (2) is a solution of (1), it follows from (5) that $\zeta_t(s, t) = \phi(\zeta(s, t))$. Thus one obtains, on placing $t = s$ in the last equation,

(7) $$\{\zeta_t(s, t)\}_{t=s} = \phi(\zeta(s, s)).$$

On placing $t = s$ in (6) and subtracting the result from (7), one obtains

(8) $$\{[\zeta(s, t) - \zeta(t, t)]_t\}_{t=s} = 0.$$

Hence, if $\epsilon > 0$ is given and s is fixed so that $t_1 < s < t_2$, finally if s' is sufficiently near s, it follows from $\{\zeta(s, t) - \zeta(t, t)\}_{t=s} = 0$ that

(9) $$|\zeta(s, s') - \zeta(s', s')| < \epsilon\,|s - s'|.$$

On the other hand, if s and t are fixed in such a way that $\zeta(s, t)$ is defined, and s' is sufficiently near s, then $\zeta(s', t)$ is defined, so that (3) and (5) imply that

(10) $$\zeta(s, t) = \psi(\zeta(s, s'), t - s'), \qquad \zeta(s', t) = \psi(\zeta(s', s'), t - s').$$

On using condition (ii) in the form (4), one obtains from (10)

(11) $$|\zeta(s, t) - \zeta(s', t)| < C\,|\zeta(s, s') - \zeta(s', s')|.$$

From (9) and (11) it is clear that

$$(12) \qquad | \zeta(s, t) - \zeta(s', t) | < \epsilon C \, | s - s' |$$

whenever $\zeta(s, t)$ is defined and s' is sufficiently near s. Since C is independent of s and s', it is seen from (12) that $\zeta_s(s, t)$ exists and is equal to 0 whenever $\zeta(s, t)$ is defined. Thus $\zeta(s; t)$ is independent of s, so that the theorem is proved.

Needless to say, the solution of (1) need not be unique if a system (2) of solutions is known satisfying (i) but not satisfying (ii).

II. *The set of split points of an ordinary differential equation on a solution.* If the vector function $\phi(\xi, t)$ of the vector ξ and the real variable t is continuous in a region R of the (ξ, t)-space, then the system

$$(1) \qquad \xi_t = \phi(\xi, t)$$

may have more than one solution passing through a given point of R. However, very little is known about the possible distribution in R of those points through which passes more than one solution of the system. The present note contains a characterization of the distribution of such points on a solution of the single equation

$$(2) \qquad x_t = f(x, t),$$

where $f(x, t)$ is a continuous function in a region R of the (x, t)-plane, and also contains a remark on the corresponding problem in case of the system (1).

It is convenient to order the points on a solution $x = g(t)$ of (1) by saying that the point P_1 is to the right of the point P_2 if P_1 belongs to a larger value of t than P_2. Then it is clear what is meant by the following definitions: The point $P \equiv (x, t)$ is called a *right split point* (a *left split point*) of (2) if two solutions of (2) exist, both containing P and such that no arc of the first solution with P as left (right) end point belongs entirely to the second solution. A point set S on a solution of (1) is said to be *right* (*left*) *closed* if any point of R which is the limit of a sequence of points on S belongs to S, provided each point of the sequence is to the right (left) of the next point. It will be proved first that:

(3) *The set of right* (*left*) *split points on any solution of* (2) *is right* (*left*) *closed.* In fact, it is well known that through any point P of R there passes a uniquely determined right upper, right lower, left upper, left lower solution of (2). It is clear that if P_1 and P_2 are on a solution α of (1) and

P_1 is to the right of P_2, then the arc P_2P_1 of α together with the right upper solution through P_1, is a right hand solution through P_2, so that the right upper solution through P_2 is never below the right hand upper solution through P_1. Similar statements hold for the other extreme solutions. Now let $\{P_n\}$ be a sequence of right split points on a solution α of (2) such that P_n is to the right of P_{n+1}, for every n, and $\{P_n\}$ converges to a point P of α. Then the right upper (lower) solution of (2) through P is never lower (higher) than the right upper (lower) solution through P_n. Thus the right hand upper and lower solutions through P_n do not coincide on any segment with P as endpoint, since the right hand upper and lower solutions through the right split point P_n do not coincide on any segment with P_n as endpoint, and any segment of α with P as left end point contains at least one point P_n. The case of left split points may be similarly treated.

(4) *If on the t-axis in the (x, t)-plane there are given point sets S_l and S_r which are left and right closed respectively, then there exists a continuous function $f(x, t)$, such that $f(0, t) = 0$ and S_l, S_r are the sets of left, right split points of* (2) *on the t-axis respectively.*

The function $f(x, t)$ will be constructed in such a way that $f(x, t) \geqq 0$ for any x and t. Then the t-axis is the right lower and left upper solution through any of its points, so that the values of $f(x, t)$ in the lower (upper) half plane alone are sufficient to decide whether any point of the t-axis is a left (right) split point or not. Thus it will be sufficient to consider the case where S_l is the empty set and $f(x, t) = 0$ whenever $x \leqq 0$ and for all t, while $f(x, t) \geqq 0$ whenever $x > 0$ and for all t.

Since the set S_r, which will from now on be denoted by S, is right closed by assumption, so that every point not in S is the left end point of an interval that does not meet S, the complement C of S on the t-axis is the sum of an at most enumerable collection of mutually disjoint intervals, the left end point of each of which is in S, while the right end point may or may not be in S. The set S itself may but need not contain intervals. In any case the set of maximal intervals contained in S is at most enumerable, and the left end point of any such interval is always contained in S. Let an enumerable set T contain the following points:

(i) The left end points of all maximal intervals in S;

(ii) The end points of all intervals constituting C;

Then any point s of S has one of the following properties:

(iii) s is contained in T;

(iv) s is interior point of a maximal interval of S.

(v) s is the limit of a monotone decreasing sequence of elements both in S and in T.

In fact, if an element s of S does not have either property (iii) or property (iv), then s is not the left end point of any interval belonging entirely to S or entirely to C. But then any interval of which s is the left end point contains the left end point of one of the intervals constituting C. Since such a left end point is contained both in S and in T, the point s has property (v).

Now let the points of T be arranged into the sequence $\{t_n\}$. Suppose that a function $f(x, t)$ has been defined on closed sets $A_1, \cdots, A_{n-1}$, corresponding to the points $t_1, \cdots, t_{n-1}$ and let there be defined, corresponding to the point t_n of T, a closed set A_n and a function $f(x, t)$ on A_n as follows:

(vi) If t_n is not the left end point of an interval either in S or in C, let A_n be the curve

$$(*) \qquad x = (t - t_n)^2/n, \qquad t_n \leqq t \leqq t_n + \tau_n,$$

where either $\tau_n = 1$ or, if the smallest number $p > 0$ such that the point $(p^2/n,\ t_n + p)$ is on one of the sets $A_1, \cdots, A_{n-1}$ exists and is not larger than 1, then $\tau_n = (n-1)p/n$. The value of $f(x, t)$ at any point of A_n is defined to be the slope of (*) at that point.

(vii) If t_n is in S and is the left end point of an interval $t_n < t < r_n$ which is contained in C, let A_n be the set defined by the inequalities

$$(**) \qquad t_n \leqq t \leqq r; \quad 0 \leqq x \leqq \min\,[(t - t_n)^2/n, 1/n],$$

while $f(x, t) = 2x/(t - t_n)$ for any point (x, t) in A_n (in particular $f(0, t_n) = 0$). Thus through any point (x, t) of A_n there passes a solution of (2) of the form $x = c(t - t_n)^2$, where $0 \leqq c \leqq 1/n$ in view of (**).

(viii) If t_n is the left end point of an interval $t_n \leqq t < r_n$ which is contained in S, let A_n be the set defined by (6), and put $f(x, t) = 2x^{\frac{1}{2}}n^{-\frac{1}{2}}$ for any point (x, t) in A_n. Thus through any point of A_n there passes a solution of (2) of the form $x = (t - c)^2/n$, where obviously $t_n \leqq c \leqq r_n$.

(ix) If t_n is in C, so that it is the left end point of an interval $t_n \leqq t < r_n$ which is contained in C, let ϵ_n be such that $0 < \epsilon_n < 1/n$, while the set A_n defined by

$$t_n \leqq t \leqq r_n; \quad 0 \leqq x \leqq \min\,[(t - t_n)/\epsilon_n, 1/n]$$

does not have a point in common with any of the sets $A_1, \cdots, A_{n-1}$ except maybe the points $(0, t_n)$ or $(0, r_n)$. Put $f(x, t) = 0$ for any point (x, t) in A_n.

Thus there may be defined by a process of complete induction a set A_n corresponding to every element t_n of T and a continuous function $f(x, t)$ on A_n. If (x, t) is any point of A_n, then clearly $x \leqq 1/n$ and $f(t, x) \leqq 2x^{\frac{1}{2}}n^{-\frac{1}{2}}$. Since the different sets A_n may have only points on the t-axis in common, it follows that, by putting $f(x, t) = 0$ for any point on the t-axis where $f(x, t)$ is not yet defined, the function $f(x, t)$ has now been defined as a continuous function on a closed set. Put $f(x, t) = 0$ for $x \leqq 0$ and for all t and let $f(x, t)$ be defined for $x \geqq 0$ as a non-negative continuous extension[3] of the function already defined. It will be shown that a point on the t-axis is a right split point of (2) if and only if the point is in S, thus completing the proof of (4) in view of the remarks immediately following (4).

Let s be any point of S. If s satisfies (iii), then it is clear from (vi), (vii), (viii) that s is a right split point of (2). Hence, if s satisfies (v), then s is a right split point of (2) in view of (3). If s satisfies (iv), then s is a split point of (2) by (viii) in view of (i). Thus any point of S is a right split point of (2). Let r be a point of C. If r is not an interior point of C, then r is not a right split point of (2) by (ix). Finally, if r is an interior point of C, then r is not a right split point of (2) by (vii) or (ix).

Let it be remarked that the notion of a split point is not as simple in case of the system (1) as in case of the equation (2). In fact it is possible to construct a function $\phi(\xi, t)$ such that a given point $P \equiv (\xi_0, t_0)$ in the (ξ, t)-space has the following properties

Any two solutions of (1) through P are identical on a t-interval containing P_0 and

If any t-interval I containing t_0 is given, then two solutions of (1) may be found which are not identical on I.

III. *On the set of split points of an ordinary differential equation on a cross line.* Let $f(x, t)$ be a continuous function in a region U of the (x, t)-plane. A simple arc α will be said to be a cross line of the differential equation

$$x_t = f(x, t) \tag{1}$$

if any solution of (1) through a point P of α has only the point P in common with α. For a sufficiently small $\epsilon > 0$, let $R_\epsilon(L_\epsilon)$ denote the set of those points

[3] It is easily seen how in the course of this extension eventual unbounded intervals contained in S or in C must be taken in account.

of U which are on solutions through points (x_0, t_0) of α and correspond to values of t satisfying $t_0 < t < t_0 + \epsilon (t_0 - \epsilon < t < t_0)$. First it will be shown that:

(2) *If, for a fixed $\epsilon > 0$, the set R_ϵ does not contain a left split point of* (1), *then the set of right split points of* (1) *on α is at most enumerable.* In fact, if P is a right split point of (1) on α, then the right upper and right lower solutions of (1) through P in R_ϵ are distinct, so that a certain open subset U_P of R_ϵ is between these two solutions. Moreover, if P and Q are distinct right split points of (1) on α, then the open sets V_p, V_q cannot have a point in common, since (1) does not have a left split point in R_ϵ, so that solutions through distinct points of α cannot have a point in common on R_ϵ. Thus (2) follows from the well-known fact that a collection of mutually disjoint open subsets of the (x, t)-plane is at most enumerable. Note that the statement of (2) remains correct if the notion "left split point" is replaced by "right split point," even if R_ϵ is not replaced by L_ϵ.

Although (2) is true, it is quite easy to construct a function $f(x, t)$, such that (1) has cross lines consisting entirely of split points. In fact it will be shown that:

(3) *If C is any nowhere dense closed set on the t-axis, then a function $f(x, t)$ exists such that both the set of all right split points and the set of all left split points of* (1) *contains all points in the (x, t)-plane of which the t-coördinate is in C.* It follows of course from (2) that the set of split points must contain more points than postulated in (3). However, the function $f(x, t)$ will be constructed in such a way that the set of split points not postulated in (3) is enumerable. It is to be noted that, according to (3), the measure of the set of split points of (1) in a bounded region may be any non-negative number less than the measure of that region. On the other hand no example is known of a function (1) for which the set of split points is dense in any region of the (x, t)-plane. The construction of the example announced in (3) will be given in several steps:

a. Functions $f(x, t)$ may be constructed such that:

(i) $f(x, t)$ is defined and continuous for every (x, t);

(ii) $f(x, t) = 0$ if either $x \leqq 0$ or $x \leqq 1$;

(iii) $|f(x, t)| < M$, where M is independent of x and t;

(iv) Any solution of (1) which is defined for all t contains at least one right split point and at least one left split point.

(v) The set of split points of (1) is enumerable.

Such a function may, for instance, be defined as follows: At any point P of any one of the curves

$$x = 4n + c(1 - \cos 2\pi t) \quad \text{or} \quad x = 4n + 2 + c(1 + \cos 2\pi t), \quad 0 \leqq t \leqq 1,$$

where c is any real number such that $-1 \leqq c \leqq 1$ and n is any positive or negative integer, put $f(x, t)$ equal to the slope at P of the curve passing through P.

b. Functions $f(x, t)$ may be constructed which have properties (i), (ii), (iii) and are such that

(vi) The set of left split points and the set of right split points on any solution of (1) which is defined for all t has the points corresponding to $t = 0$ and to $t = 1$ as cluster points;

(vii) Although, according to (ii), (vi) and (3) of Section II, every point $(x, 0)$ is a right split point of (1) and every point $(x, 1)$ is a left split point of (1), the set of split points for which $0 < t < 1$ is enumerable.

Such a function may, for instance, be defined as follows:

If $t \leqq 0$ or $t \geqq 0$, put $f(x, t) = 0$;
if $2^{-n} \leqq t \leqq 2^{1-n}$, put $f(x, t) = n^{-1} g(n 2^n x, 2^n t - 1)$;
finally, if $1 - 2^{1-n} \leqq t \leqq 1 - 2^{-n}$, put $f(x, t) = n^{-1} g(n 2^n x, 2^n (t - 1) + 2)$.

Here $g(x, t)$ is a function with properties (i), (ii), (iii), (iv) and (v) and n is any integer larger than one. That the function $f(x, t)$ thus defined has the required properties may be easily verified. A function with properties (i), (ii), (iii), (vi) and (vii) will be denoted by $h(x, t)$.

c. Now let C be a given nowhere dense closed set on the t-axis, and let the intervals constituting the complement of C be [4] $a_n < t < b_n$, $n = 1, 2, \cdots$.

A function $f(x, t)$, satisfying the requirements of (3) may now be defined as follows:

[4] It is easy to see how an eventual unbounded interval in C should be taken in account.

If t is in C, put $f(x, t) = 0$;
if $a_n \leqq t \leqq b_n$, for some n, put $f(x, t) = \frac{1}{n} h\left(\frac{nx}{b_n - a_n}, \frac{t - a_n}{b_n - a_n}\right)$.

The function $f(x, t)$ thus defined clearly is continuous for every x and t, since $h(x, t)$ has properties (i), (ii), (iii). Let t_0 be an element of C. Since C is nowhere dense, either there exists an n such that $t_0 = a_n < b_n$ or, for every $\epsilon > 0$, there exists an n such that $t_0 < a_n < b_n < t_0 + \epsilon$. Thus it follows from (vi) and from (3) of Section II that a point $P \equiv (x_0, t_0)$ is a left split point of (1) if t_0 is in C. It is similarly seen that P also is a right split point of (1). Finally, it is clear from (vii) that the set of those split points of (1) whose t-coördinate is not in C is an enumerable set.

THE JOHNS HOPKINS UNIVERSITY.

ON A SYMMETRICAL CANONICAL REDUCTION OF THE PROBLEM OF THREE BODIES.[1]

By E. R. van Kampen and Aurel Wintner.

Introduction. It is known [2] that the classical first integrals of the problem of three bodies permit reducing the degree of freedom from 9 to 4 in a canonical form and that on the basis of these integrals no further reduction of the general problem is possible, if the independent variable of the reduced system is the same as that of the original system, i. e., if no use is made of the energy integral. In the particular case of the planar problem of three bodies the reduced degree of freedom is 3 instead of 4. In the latter case Murnaghan [3] recently gave a symmetric treatment of the problem by using the 3 mutual distances as the 3 coördinates of the reduced system.

The object of this paper is to extend this symmetrical reduction to the non-planar case by using as the 4 coördinates the 3 mutual distances and an angular variable which is symmetrical with respect to the three bodies and becomes an ignorable coördinate in the planar case. Correspondingly, the treatment given by Murnaghan [4] for the planar case is based on a "modification," in the sense of Routh, of the Lagrangian function of the original problem. The treatment to be given in the present paper both for the planar and non-planar case is based on the consideration of the Hamiltonian instead of the Lagrangian equations. Actually, the reduction will be composed of two canonical transformations (Sections II and III) which are independent of the form of the Hamiltonian function. On applying these canonical transformations to the Hamiltonian system of the problem of three bodies (Sections IV and V), the desired reduced equations immediately follow.

It is suggested by the simplicity and symmetry of the result that the reduced equations can be applied with success to several classical problems concerning particular solutions of a preassigned nature.

I. *Extended canonical transformations and invariant relations.* The

[1] Received November 9, 1936.

[2] For the extensive literature of the subject, cf. R. Marcolongo, *Il problema dei tre corpi*, Milano, 1919, Chap. 2, pp. 27-41.

[3] F. D. Murnaghan, "A symmetric reduction of the planar three-body problem," *American Journal of Mathematics*, vol. 58 (1936), pp. 829-832.

[4] *Loc. cit.*, [3], p. 831.

object of this Section I is to collect for later use some rules dealing with the introduction of new variables into a canonical system

$$(1) \qquad q'_l = \partial H/\partial p_l, \qquad p'_l = -\partial H/\partial q_l; \qquad (l = 1, 2, \cdots, n)$$

with n degrees of freedom, where the prime denotes differentiation with respect to t and the Hamiltonian function

$$(2) \qquad H = H(q_1, \cdots, q_n, p_1, \cdots, p_n)$$

is supposed to have continuous partial derivatives of the second order in the $2n$-dimensional region under consideration.

(i) Let κ_λ and π_λ, where $\lambda = 1, \cdots, \nu$, be the coördinates and impulses in a 2ν-dimensional phase-space, where ν is less than the degree of freedom, n, of (1). Put

$$(3_1) \qquad \kappa_\lambda = \phi_\lambda(q_1, \cdots, q_n); \qquad (\lambda = 1, 2, \cdots, \nu),$$

$$(3_2) \qquad p_l = \sum_{\lambda=1}^{\nu} \pi_\lambda \, \partial\phi_\lambda/\partial q_l; \qquad (l = 1, 2, \cdots, n),$$

and restrict the n coördinates q_l by $n - \nu$ relations

$$(4) \qquad \psi_\mu(q_1, \cdots, q_n) = 0; \qquad (\mu = 1, 2, \cdots, n - \nu).$$

It will be asssumed that the n functions $\phi_1, \cdots, \phi_\nu, \psi_1, \cdots, \psi_{n-\nu}$ of the n variables $q_1, \cdots, q_n$ have in the n-dimensional q-region under consideration continuous partial derivatives of the third order and a non-vanishing Jacobian. Thus from (3_1) and (4)

$$(5) \qquad q_l = f_l(\kappa_1, \cdots, \kappa_\nu); \qquad (l = 1, 2, \cdots, n),$$

where the functions f_l are locally univalued and have continuous partial derivatives of the third order. On substituting (3_2) and (5) into (2), one obtains a function of the κ_λ and π_λ which will be denoted by H, so that, in virtue of (3_2) and (5),

$$(6) \qquad H = \mathrm{H}, \text{ where } \mathrm{H} = \mathrm{H}(\kappa_1, \cdots, \kappa_\nu, \pi_1, \cdots, \pi_\nu).$$

Now the function H thus defined has the following property: For those solutions of (1) along which (4) is satisfied the canonical system (1) is equivalent to the canonical system

$$(7) \qquad \kappa'_\lambda = \partial\mathrm{H}/\partial\pi_\lambda, \qquad \pi'_\lambda = -\partial\mathrm{H}/\partial\kappa_\lambda; \qquad (\lambda = 1, 2, \cdots, \nu)$$

of ν $(< n)$ degrees of freedom. The proof requires but straightforward differentiations and substitutions.

(ii) Let again κ_λ and π_λ, where $\lambda = 1, 2, \cdots, \nu$ and $\nu < n$, be the coördinates and impulses in a 2ν-dimensional phase-space, but replace (3_1), (3_2) by

$$(8_1) \qquad q_l = F_l(\kappa_1, \cdots, \kappa_\nu); \qquad (l = 1, 2, \cdots, n),$$

$$(8_2) \qquad \pi_\lambda = \sum_{l=1}^{n} p_l \, \partial F_l / \partial \kappa_\lambda; \qquad (\lambda = 1, 2, \cdots, \nu),$$

and (4) by

$$(9) \qquad \sum_{l=1}^{n} p_l G_{l\mu}(\kappa_1, \cdots, \kappa_\nu) = 0; \qquad (\mu = 1, 2, \cdots, n - \nu).$$

It will be assumed that, in the ν-dimensional κ-region under consideration, the n functions F_l and the $n(n-\nu)$ functions $G_{l\mu}$ have continuous partial derivatives of the third order and that, in the same region, the determinant of the $\nu + (n - \nu) = n$ linear forms (8_2), (9) in the n impulses p_l is distinct from zero. Thus, from (8_2) and (9),

$$(10) \qquad p_l = \sum_{\lambda=1}^{\nu} \pi_\lambda \Gamma_{l\lambda}(\kappa_1, \cdots, \kappa_\nu); \qquad (l = 1, 2, \cdots, n),$$

where the $n\nu$ functions $\Gamma_{l\mu}$ are uniquely determined and have continuous partial derivatives of the second order. On substituting (8_1) and (10) into (2), one obtains a function H of the κ_λ and π_λ, so that (6) holds in virtue of (8_1) and (10). Now the function H thus defined has the following property: For those solutions of (1) along which (9) is satisfied the canonical system (1) is equivalent to the canonical system (7). The proof requires, as in the case (i), but straightforward differentiations and substitutions,[5] as seen by differentiating (8_1) and (10) with respect to t and (6) with respect to κ_ν, π_ν.

(iii) Let $i = n - m + 1, \; n - m + 2, \cdots, n$, where m is a positive integer less than the degree of freedom of a canonical system (1), and let c_i be a constant and h_i a function of $q_1, \cdots, q_{n-m}$ and $p_1, \cdots, p_{n-m}$ which has continuous partial derivatives of the second order in the region under consideration. Suppose that the m pairs of relations

$$(11) \quad q_i = h_i(q_1, \cdots, q_{n-m}, p_1, \cdots, p_{n-m}), \;\; p_i = c_i; \;\; (i = n - m + 1, \cdots, n)$$

[5] Cf. also T. Levi-Civita, "Solla introduzione di vincoli olonomi nelle equazioni dinamiche di Hamilton," *Atti del Reale Istituto Veneto*, ser. 8, vol. 18_1 (1916), pp. 387-395.

form an invariant system[6] of (1). On substituting (11) into (2), one obtains a function K of $q_1, \cdots, q_m, p_1, \cdots, p_m$ and the c_i, so that, on considering the numbers c_i as fixed,

$$(12) \qquad H = K, \text{ where } K = K(q_1, \cdots, q_m, p_1, \cdots, p_m).$$

Now the function K thus defined has the following property: For those solutions of (1) along which (11) is satisfied the canonical system (1) is equivalent to the canonical system

$$(13) \qquad q'_k = \partial K/\partial p_k, \qquad p'_k = -\partial K/\partial q_k; \qquad (k = 1, 2, \cdots, n-m).$$

This is easily verified by noting that

$$\partial H/\partial q_i = 0; \qquad (i = n-m+1, \cdots, n)$$

in virtue of the invariant system (11) of (1).

All three rules (i), (ii), (iii) are to the effect that under the conditions specified the new Hamiltonian function, H or K, is obtained from the original Hamiltonian function H by actual substitution of the transformation formulae. In such a case one says that the transformation formulae define a *completely canonical* transformation. It may be mentioned that all three rules (i), (ii), (iii) are very special cases of a general theorem dealing with canonical systems which are subjected either to constraints or to invariant relations.[7]

A notation. From now on i, j, k will denote subscripts each of which has one of the three values 1, 2, 3. The summation symbol $\Sigma\, a_{ijk}$ will denote the sum

$$\Sigma\, a_{ijk} = a_{123} + a_{231} + a_{312} \qquad (\neq \Sigma\, a_{ikj}),$$

so that, in particular

$$\Sigma\, a_{ij} = a_{12} + a_{23} + a_{31}, \qquad \Sigma\, a_i = a_1 + a_2 + a_3.$$

[6] A system of s relations

$$(*) \qquad r_h(x_1, \cdots, x_k, t) = 0; \qquad (h = 1, \cdots, s)$$

is an invariant system of a system of k differential equations

$$(**) \qquad dx_m/dt = X_m(x_1, \cdots, x_k, t); \qquad (m = 1, \cdots, k),$$

if the functions r_h have continuous partial derivatives of the first order and

$$dr_h(x_1, \cdots, x_k, t)/dt = 0; \qquad (h = 1, \cdots, s)$$

is a consequence of (*) and (**).

[7] Cf. a forthcoming paper where holonomic and certain non-holonomic constraints will be treated by a method used in the case of free canonical systems by E. R. van Kampen and A. Wintner, "On the canonical transformations of Hamiltonian systems," *American Journal of Mathematics*," vol. 58 (1936), pp. 851-863.

An equation of the type $a_{ijk} = b_{ijk}$ will be meant to hold not necessarily for all 6, merely for the 3 cyclic, permutations $(1,2,3)$, $(2,3,1)$, $(3,1,2)$ of $(i,j,k) = (1,2,3)$.

II. *A canonical transformation of a 16-dimensional phase-space.* Let P_0 be an 18-dimensional phase-space of 9 coördinates x_i, y_i, z_i and of 9 impulses X_i, Y_i, Z_i, where $i = 1, 2, 3$. Consider that subspace of P_0 on which

$$(14_1)\quad \begin{vmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{vmatrix} = 0; \qquad (14_2)\quad \Sigma \frac{1}{m_i} \begin{vmatrix} X_i & Y_i & Z_i \\ x_j & y_j & z_j \\ x_k & y_k & z_k \end{vmatrix} = 0,$$

where m_1, m_2, m_3 are three positive constants. Suppose further that

$$(15_1)\quad r = 2; \qquad (15_2)\quad \Sigma z_i^2 \neq 0,$$

where r denotes the rank of the matrix of (14_1). The subspace of P_0 characterized by the four conditions (14_1), (14_2), (15_1), (15_2) will be denoted by P, so that P is 16-dimensional.

Let Π denote the 16-dimensional phase-space of 8 coördinates ν, ι; ξ_i, η_i such that

$$(16_1)\quad \Delta \neq 0; \qquad (16_2)\quad \sin \iota \neq 0$$

and of 8 impulses N, I; Ξ_i, H_i, where $i = 1, 2, 3$ and Δ is an abbreviation for

$$(17)\qquad \Delta = \Sigma\, m_i m_j (\xi_i \eta_j - \xi_j \eta_i)^2.$$

One can express condition (16_1) without reference to the three constants $m_i > 0$ by requiring that the two vectors (ξ_1, ξ_2, ξ_3), (η_1, η_2, η_3) be linearly independent. It is clear from (17) that

$$(18)\qquad \Delta = M_{\xi\xi} M_{\eta\eta} - M_{\xi\eta}^2,$$

where

$$(19)\qquad M_{\xi\xi} = \Sigma\, m_i \xi_i^2, \qquad M_{\xi\eta} = \Sigma\, m_i \xi_i \eta_i, \qquad M_{\eta\eta} = \Sigma\, m_i \eta_i^2.$$

Now put

$$(20_1)\qquad \begin{aligned} x_i &= \xi_i \cos \nu - \eta_i \sin \nu \cos \iota, \\ y_i &= \xi_i \sin \nu + \eta_i \cos \nu \cos \iota, \\ z_i &= \xi_i \cdot 0 \quad + \eta_i \sin \iota, \end{aligned}$$

and

$$(20_2)\qquad \begin{aligned} N &= \Sigma\, \xi_i (-X_i \sin \nu + Y_i \cos \nu) - \cos \iota\, \Sigma\, \eta_i (X_i \cos \nu + Y_i \sin \nu), \\ I &= \Sigma\, \eta_i (X_i \sin \nu \sin \iota - Y_i \cos \nu \sin \iota + Z_i \cos \iota), \end{aligned}$$

finally

$$(20_3)\qquad \begin{aligned} \Xi_i &= X_i \cos\nu + Y_i \sin\nu,\\ \mathrm{H}_i &= -X_i \sin\nu\cos\iota + Y_i\cos\nu\cos\iota + Z_i\sin\iota, \end{aligned}$$

where $i=1,2,3$. It will be shown that the formulae (20_1), (20_2), (20_3) define a completely canonical transformation of the two 16-dimensional phase-spaces P, Π into each other.

First, if G_x, G_y, G_z and Γ are defined by

$$(21)\qquad G_x = \Sigma(y_iZ_i - z_iY_i),\quad G_y = \Sigma(z_iX_i - x_iZ_i),\quad G_z = \Sigma(x_iY_i - y_iX_i)$$

and

$$(22)\qquad \Gamma = \Sigma(\xi_i\mathrm{H}_i - \eta_i\Xi_i),$$

then

$$(23_1)\qquad \mathrm{N} = G_z,\qquad \mathrm{I} = G_x\cos\nu + G_y\sin\nu,$$

$$(23_2)\qquad \Gamma = G_x\sin\nu\sin\iota - G_y\cos\nu\sin\iota + G_z\cos\iota$$

are identities in virtue of (20_1), (20_2), (20_3). This is seen by expressing G_x, G_y, G_z and Γ by means of (20_1), (21) and (20_3), (22) in terms of $\nu, \iota;\ \xi_i, \eta_i;\ X_i, Y_i, Z_i$ and substituting the result and (20_2) into (23_1), (23_2).

Next, on placing

$$(24)\qquad \mathrm{T}_i = X_i\sin\nu\sin\iota - Y_i\cos\nu\sin\iota + Z_i\cos\iota,$$

where $i=1,2,3$, it is clear that (20_3) and (24) determine, for every fixed i, an orthogonal transformation of the vector (X_i, Y_i, Z_i) into the vector $(\Xi_i, \mathrm{H}_i, \mathrm{T}_i)$, so that, the inverse transformation being obtained by transposing the matrix,

$$(25)\qquad \begin{aligned} X_i &= \Xi_i\cos\nu - \mathrm{H}_i\sin\nu\cos\iota + \mathrm{T}_i\sin\nu\sin\iota,\\ Y_i &= \Xi_i\sin\nu + \mathrm{H}_i\cos\nu\cos\iota - \mathrm{T}_i\cos\nu\sin\iota,\\ Z_i &= \Xi_i\cdot 0 \quad + \mathrm{H}_i\sin\iota \qquad + \mathrm{T}_i\cos\iota. \end{aligned}$$

Furthermore,

$$(26_1)\qquad \Sigma\,\eta_i\mathrm{T}_i = \mathrm{I};\qquad \Sigma\,\xi_i\mathrm{T}_i = \frac{\Gamma\cos\iota - \mathrm{N}}{\sin\iota},\ \text{if } \sin\iota\neq 0$$

are identities, while

$$(26_2)\qquad \Sigma\, m_jm_k(\xi_j\eta_k - \xi_k\eta_j)\mathrm{T}_i = 0$$

is an identity in virtue of (14_2). In fact, the first of the relations (26_1) is obvious from the second of the relations (20_2) and from (24), while the second of the relations (26_1) follows by substituting (20_3) into (22) and using (20_1) and (24). Finally, on substituting (20_1) and (25) into the determinants occurring in (14_2), a straightforward calculation shows that (14_2) may be written in the form (26_2).

Now it is easily verified by direct substitution that (14_1), (15_1) and (15_2) are implied by (20_1), (16_1) and (16_2). Conversely, if 9 given values x_i, y_i, z_i, where $i=1,2,3$, satisfy (14_1), (15_1) and (15_2), then the equations (20_1), (16_1) and (16_2) may be satisfied by 8 values $\nu, \iota, \xi_i, \eta_i$ and, on disregarding the trigonometrical ambiguity of the angles ν and ι, the 8 values $\nu, \iota, \xi_i, \eta_i$ thus obtained are uniquely determined by the 9 values x_i, y_i, z_i. This is seen from the fact that the coefficients of the linear substitution (20_1) form the first and second columns of a three-rowed orthogonal matrix.

Accordingly, conditions (14_1), (15_1) and (15_2) are equivalent to conditions (16_1) and (16_2) in virtue of (20_1) alone. On the other hand, on identifying (20_1) with (8_1), it is seen by partial differentiation of (20_1) that the system (20_2), (20_3) goes over into (8_2). Hence, in order to conclude from the rule (ii) of Section I that (20_1), (20_2), (20_3) determine a completely canonical transformation of the two 16-dimensional phase-spaces P, Π into each other, it is sufficient to prove that the 9 linear relations (14_2), (20_2), (20_3) for the 9 variables X_i, Y_i, Z_i have a non-vanishing determinant. Now it is seen from (24) and (25) that this 9-rowed determinant cannot vanish, unless so does the 3-rowed determinant of the 3 linear conditions (26_1), (26_2) for the 3 variables T_i. Finally, this 3-rowed determinant is easily verified to be identical with the expression (17) and is, therefore, distinct from zero in view of the assumption (16_1).

Let it be noted for later application that

$$(27)\qquad \Sigma\, m_i^{-1}(X_i^2+Y_i^2+Z_i^2)=\Sigma\, m_i^{-1}(\Xi_i^2+\mathrm{H}_i^2)+\Sigma\, m_i^{-1}\mathrm{T}_i^2$$

and

$$(28)\qquad \Sigma\, m_i^{-1}\mathrm{T}_i^2\Delta=\mathrm{I}^2\mathrm{M}_{\xi\xi}-2\mathrm{I}\frac{\Gamma\cos\iota-\mathrm{N}}{\sin\iota}\mathrm{M}_{\xi\eta}+\frac{(\Gamma\cos\iota-\mathrm{N})^2}{\sin^2\iota}\mathrm{M}_{\eta\eta}.$$

In fact, since (25) is an orthogonal substitution, (27) is obvious. On the other hand, on solving the 3 linear equations (26_1), (26_2) with respect to $\mathrm{T}_1, \mathrm{T}_2, \mathrm{T}_3$ and using the definitions (17) and (19), one obtains

$$m_i^{-1}\mathrm{T}_i\Delta=\mathrm{I}(\eta_i\mathrm{M}_{\xi\xi}-\xi_i\mathrm{M}_{\xi\eta})-\frac{\Gamma\cos\iota-\mathrm{N}}{\sin\iota}(\eta_i\mathrm{M}_{\xi\eta}-\xi_i\mathrm{M}_{\eta\eta}).$$

Hence (28) follows by direct substitution, if one uses (18) and (19).

III. *A canonical transformation of an 8-dimensional phase-space.* Let Ψ_0 denote the 12-dimensional phase-space of 6 coördinates ξ_i, η_i and of 6 impulses Ξ_i, H_i, where $i=1,2,3$. For 3 given positive constants m_i, let Ψ denote that subspace of Ψ_0 on which

$$(29_1) \qquad \Sigma\, m_i\xi_i = 0, \qquad \Sigma\, m_i\eta_i = 0;$$

$$(29_2) \qquad \Sigma\, \Xi_i = 0, \qquad \Sigma\, \mathrm{H}_i = 0;$$

$$(29_3) \qquad \delta > 0, \text{ where } \delta = \tfrac{1}{2}\begin{vmatrix} \xi_1 & \eta_1 & 1 \\ \xi_2 & \eta_2 & 1 \\ \xi_3 & \eta_3 & 1 \end{vmatrix},$$

so that the space Ψ is 8-dimensional. Condition (29_3) may be expressed by saying that the 3 points (ξ_i, η_i) form in a (ξ, η)-plane a positively oriented triangle of area $\delta \neq 0$. Let ϕ_i denote the inclination of that oriented side of this triangle which is opposite to the vertex (ξ_i, η_i), so that

$$(30) \quad \cos\phi_i = \frac{\xi_k - \xi_j}{[(\xi_k - \xi_j)^2 + (\eta_k - \eta_j)^2]^{\frac{1}{2}}}, \quad \sin\phi_i = \frac{\eta_k - \eta_j}{[(\xi_k - \xi_j)^2 + (\eta_k - \eta_j)^2]^{\frac{1}{2}}};$$

where $[\ ]^{\frac{1}{2}} > 0$. Furthermore,

$$(31) \qquad \theta_i = \phi_k - \phi_j,$$

where θ_i denotes the oriented exterior angle at the vertex (ξ_i, η_i) of the triangle, so that

$$(32) \qquad \Sigma\, \theta_i = 0$$

and not only $\Sigma\, \theta_i \equiv 0 \pmod{2\pi}$. The notations (17), (19) of Section II will be used in the space Ψ also. On multiplying the i-th row of the determinant (29_3) by $m_i^{\frac{1}{2}}$ and calculating the square of the resulting determinant by column-by-column multiplication, it is easily seen from (18), (19) and (29_1) that

$$(33) \qquad \Delta = 4\mu\delta^2,$$

where μ is the positive constant

$$(34) \qquad \mu = m_1m_2m_3/(m_1 + m_2 + m_3).$$

Let Θ denote the 8-dimensional phase-space of 4 coördinates ω, ρ_i such that

$$(35) \qquad 0 < \rho_i < \rho_j + \rho_k$$

and of 4 impulses Ω, P_i, where $i = 1, 2, 3$. Condition (35) means that the ρ_i are the sides of a non-degenerate triangle.

Now it will be shown that the formulae

$$(36_1) \qquad \omega = \tfrac{1}{3}\,\Sigma\,\phi_i,$$

$$(36_2) \qquad \rho_i = [(\xi_k - \xi_j)^2 + (\eta_k - \eta_j)^2]^{\frac{1}{2}},$$

$$(36_3)\qquad \begin{aligned} \Xi_i &= \mathrm{P}_j \cos\phi_j - \mathrm{P}_k \cos\phi_k - \frac{\Omega}{3}\left(\frac{\sin\phi_j}{\rho_j} - \frac{\sin\phi_k}{\rho_k}\right),\\ \mathrm{H}_i &= \mathrm{P}_j \sin\phi_j - \mathrm{P}_k \sin\phi_k + \frac{\Omega}{3}\left(\frac{\cos\phi_j}{\rho_j} - \frac{\cos\phi_k}{\rho_k}\right), \end{aligned}$$

define in virtue of (30) a completely canonical transformation of the two 8-dimensional phase-spaces Ψ, Θ into each other.

First, (36_2) means the identification of the two triangles mentioned above. Hence

$$(37)\qquad \cos\theta_i = \frac{\rho_i^2 - \rho_j^2 - \rho_k^2}{2\rho_j\rho_k}, \qquad \sin\theta_i = \frac{2\delta}{\rho_j\rho_k},$$

where δ is the area (29_3), so that

$$(38)\qquad 16\delta^2 = (\rho_1+\rho_2+\rho_3)(-\rho_1+\rho_2+\rho_3)(\rho_1-\rho_2+\rho_3)(\rho_1+\rho_2-\rho)\,;\ \delta > 0.$$

Furthermore, from (36_1) and (31),

$$(39)\qquad \phi_i = \omega + \tfrac{1}{3}(\theta_j - \theta_k),$$

while from (36_2) and (30)

$$(40)\qquad \xi_k - \xi_j = \rho_i \cos\phi_i, \qquad \eta_k - \eta_j = \rho_i \sin\phi_i.$$

It is seen from (40) and (29_1) that

$$(41)\qquad \begin{aligned} (m_1+m_2+m_3)\xi_i &= m_k\rho_j \cos\phi_j - m_j\rho_k \cos\phi_k\\ (m_1+m_2+m_3)\eta_i &= m_k\rho_j \sin\phi_j - m_j\rho_k \sin\phi_k. \end{aligned}$$

Now (39) represents in view of (37) and (38) the 3 angles ϕ_i in terms of the 4 coördinates ω, ρ_i. Thus (41) represents the 6 coördinates ξ_i, η_i of the phase-space Ψ in terms of the 4 coördinates ω, ρ_i of the phase-space Θ. Accordingly, on identifying the system (36_1), (36_2) with (3_1) and (29_1) with (4), it is clear that the Jacobian mentioned under (i) in Section I is distinct from zero. Finally, partial differentiation of (36_1), (36_2) shows that (36_3) goes over into (3_2) in view of (30). This proves that (36_1), (36_2), (36_3) define a completely canonical transformation. Incidentally, (36_3) clearly implies (29_2).

Let it be mentioned for later application that

$$(42)\qquad \begin{aligned} \Sigma\, m_i^{-1}(\Xi_i^2 + \mathrm{H}_i^2) &= \Sigma\, m_i^{-1}(\mathrm{P}_j^2 + \mathrm{P}_k^2 - 2\mathrm{P}_j\mathrm{P}_k\cos\theta_i)\\ &\quad + \frac{2\Omega}{3}\Sigma\, m_i^{-1}\left(\frac{\mathrm{P}_j}{\rho_k} - \frac{\mathrm{P}_k}{\rho_j}\right)\sin\theta_i\\ &\quad + \frac{\Omega^2}{9}\Sigma\, m_i^{-1}\left(\frac{1}{\rho_j^2} + \frac{1}{\rho_k^2} - \frac{2\cos\theta_i}{\rho_j\rho_k}\right). \end{aligned}$$

In fact, on calculating $\Xi_i^2 + \mathrm{H}_i^2$ from (36_3) and using then (31), the identity (42) follows. Similarly,

$$\Gamma = \Omega, \tag{43}$$

where Γ is defined by (22). In fact, on substituting (36_3) into (22) and using then (40), one clearly obtains (43). Finally,

$$\mathrm{M}_{\eta\eta} = \mu \sum m_i^{-1} \rho_i^2 \sin^2 [\omega + \tfrac{1}{3}(\theta_k - \theta_j)]. \tag{44}$$

In fact, since

$$\sum m_i \eta_i^2 = \mu \sum m_i^{-1} (\eta_k - \eta_j)^2$$

is an identity in virtue of (29_1) and (34), it is seen from (39) and (40) that (44) is equivalent to the definition (19) of $\mathrm{M}_{\eta\eta}$. Similar formulae hold for $\mathrm{M}_{\xi\xi}$ and $\mathrm{M}_{\xi\eta}$.

IV. *Application to the non-planar problem of three bodies.* The completely canonical substitutions given in Sections II and III may be applied to a symmetrical reduction of the degree of freedom of the non-planar problem of three bodies from 9 to 4, using merely the conservation of the center of mass and of the angular momentum.

Let x_i, y_i, z_i be the Cartesian coördinates of the mass m_i in a barycentric inertial coördinate system (x, y, z), and let the plane $z = 0$ be chosen such that its positive normal has the direction of the vector which represents the constant angular momentum, this vector being not the zero vector in view of the assumption that the motion is non-planar. If C denotes the length of this vector and (X_i, Y_i, Z_i) the impulse vector $(m_i x'_i, m_i y'_i, m_i z'_i)$ of m_i, the conservation of the angular momentum is represented by the invariant relations

$$G_x = 0, \quad G_y = 0, \quad G_z = C, \tag{45}$$

where G_x, G_y, G_z are defined by (21). The preservation of the center of mass is expressed by the invariant relations

$$\sum m_i x_i = 0, \qquad \sum m_i y_i = 0, \qquad \sum m_i z_i = 0; \tag{46_1}$$

$$\sum X_i = 0, \qquad \sum Y_i = 0, \qquad \sum Z_i = 0, \tag{46_2}$$

since (x, y, z) is a barycentric inertial coördinate system. Finally, the Hamiltonian function is

$$H = \tfrac{1}{2} \sum m_i^{-1} (X_i^2 + Y_i^2 + Z_i^2) - \sum m_j m_k \rho_i^{-1}, \tag{47}$$

where

$$\rho_i = [(x_k - x_j)^2 + (y_k - y_j)^2 + (z_k - z_j)]^{\frac{1}{2}}. \tag{48}$$

Since it is assumed that the motion is non-planar, and since the equations of motion define the coördinates and impulses as analytic functions of the time t, it is clear [8] that those values of t for which the three masses are either collinear or lie in the invariable plane $z = 0$ have no cluster point on the t-axis and can, therefore, be excluded from the considerations of this Section IV. Thus conditions (15_1), (15_2) are satisfied. Furthermore, (14_1) and (14_2) are implied by (46_1) and (46_2), since $m_i > 0$.

Now introduce in place of the barycentric inertial coördinate system (x, y, z) a barycentric but non-inertial coördinate system (ξ, η, ζ) in such a way that the plane $\zeta = 0$ is, for every t, the plane of the triangle formed by the three masses and that the ξ-axis is in the invariable plane $z = 0$ for every t. Then, if ν denotes the node and ι the inclination of the moving plane $\zeta = 0$ with reference to the fixed plane $z = 0$, the Eulerian representation of a rotation shows that the transformation formulae of the coördinates are given by (20_1). Furthermore, (16_1) and (16_2) are satisfied, since the three masses are neither collinear nor all contained in the (x, y)-plane. Now (20_1), (20_2), (20_3) were shown (Section II) to define a completely canonical transformation. Hence, on expressing the Hamiltonian function (47) by means of the transformation formulae in terms of the coördinates ν, ι, ξ_i, η_i and the canonically conjugate variables N, I, Ξ_i, H_i, and denoting the resulting function again by H, the equations of motion are

$$(49_1) \quad \nu' = \frac{\partial H}{\partial \mathrm{N}}, \quad \mathrm{N}' = -\frac{\partial H}{\partial \nu}; \quad \iota' = \frac{\partial H}{\partial \mathrm{I}}, \quad \mathrm{I}' = -\frac{\partial H}{\partial \iota},$$

$$(49_2) \quad \xi'_i = \frac{\partial H}{\partial \Xi_i}, \quad \Xi'_i = -\frac{\partial H}{\partial \xi_i}; \quad \eta'_i = \frac{\partial H}{\partial \mathrm{H}_i}, \quad \mathrm{H}'_i = -\frac{\partial H}{\partial \eta_i},$$

where

$$(50) \quad H = \tfrac{1}{2}\Sigma\, m_i^{-1}(\Xi_i^2 + \mathrm{H}_i^2) - \Sigma\, m_j m_k \rho_i^{-1} + \frac{1}{2\Delta}\left[\mathrm{I}^2 \mathrm{M}_{\xi\xi} - 2\mathrm{I}\,\frac{\Gamma\cos\iota - \mathrm{N}}{\sin\iota}\,\mathrm{M}_{\xi\eta} + \frac{(\Gamma\cos\iota - \mathrm{N})^2}{\sin^2\iota}\,\mathrm{M}_{\eta\eta}\right]$$

in view of (47), (27), (28). It is understood that in (50)

$$\Gamma,\ \Delta,\ \mathrm{M}_{\xi\xi},\ \mathrm{M}_{\xi\eta},\ \mathrm{M}_{\eta\eta} \text{ and } \rho_i$$

are thought of as expressed by means of (22), (17), (19) and (36_2) in terms of ξ_i, η_i, Ξ_i, H_i.

In order to reduce the degree of freedom of the simultaneous system (49_1), (49_2) from 8 to 6 by means of the invariant relations (45), one can separate

[8] In fact, it is well known that if a solution is collinear, then it is necessarily planar. Cf.[12]

the 4 equations (49_1) from the 12 equations (49_2) as follows: First, from (23_1), (23_2) and (45),

$$(51) \qquad \mathrm{N} = C, \qquad \mathrm{I} = 0, \qquad \cos \iota = \Gamma / C,$$

while the fourth of the variables belonging to (49_1), namely the node ν, does not occur in the Hamiltonian function (50). This fact and (51) represent an invariant system of $2m = 4$ relations of the type (11). Thus it is seen from the rule (iii) of Section I that (49_2) may be replaced by

$$(52) \qquad \xi'_i = \frac{\partial K}{\partial \Xi_i}, \quad \Xi'_i = -\frac{\partial K}{\partial \xi_i}; \qquad \eta'_i = \frac{\partial K}{\partial \mathrm{H}_i}, \quad \mathrm{H}'_i = -\frac{\partial K}{\partial \eta_i},$$

where the Hamiltonian function

$$(53_1) \qquad K = K(\xi_1, \cdots, \eta_3, \Xi_1, \cdots, \mathrm{H}_3)$$

is obtained from (50) by substituting (51) into (50). Consequently,[9]

$$(53_2) \qquad K = \tfrac{1}{2} \Sigma\, m_i^{-1} (\Xi_i^2 + \mathrm{H}_i^2) - \Sigma\, m_j m_k \rho_i^{-1} + \tfrac{1}{2} (C^2 - \Gamma^2) \mathrm{M}_{\eta\eta} / \Delta,$$

where Γ, Δ, $\mathrm{M}_{\xi\eta}$ and ρ_i are thought of as expressed by means of (22), (17), (19) and (36_2) in terms of ξ_i, η_i, Ξ_i, H_i. The conservative canonical system (52) with 6 degrees of freedom possesses the 4 invariant relations (29_1), (29_2), as easily verified either from (20_1), (20_2) and (46_1), (46_2) or by partial differentiation of (53_2).

Now the degree of freedom of the system (52) may be reduced by means of the invariant relations (29_1), (29_2) from 6 to 4 by using the canonical transformation treated in Section III. In fact, (29_3) is satisfied in view of the assumption that the three masses m_i are not colinear. Hence the result of Section III is applicable and shows that (52) may be replaced by the 4 pairs of simultaneous equations

$$(54_1) \qquad \omega' = \frac{\partial \mathrm{K}}{\partial \Omega}, \qquad \Omega' = -\frac{\partial \mathrm{K}}{\partial \omega},$$

$$(54_2) \qquad \rho'_i = \frac{\partial \mathrm{K}}{\partial \mathrm{P}_i}, \qquad \mathrm{P}'_i = -\frac{\partial \mathrm{K}}{\partial \rho_i}, \qquad (i = 1, 2, 3),$$

where the Hamiltonian function

$$(55_1) \qquad \mathrm{K} = \mathrm{K}(\omega, \rho_1, \rho_2, \rho_3, \Omega, \mathrm{P}_1, \mathrm{P}_2, \mathrm{P}_3)$$

[9] This agrees with a result of Levi-Civita, "Sulla riduzione del problema dei tre corpi," *Atti del Reale Istituto Veneto*, ser. 8, vol. 17_1 (1915), pp. 907-939, more particularly p. 923, formula (28).

is obtained by expressing the Hamiltonian function (53_1) by means of (36_1), (36_2), (36_3) and (30) in terms of ω, ρ_i, Ω, P_i, so that

$$(55_2)\quad \mathrm{K}=\Sigma\frac{1}{2m_i}(\mathrm{P}_j{}^2+\mathrm{P}_k{}^2-2\mathrm{P}_j\mathrm{P}_k\cos\theta_i)$$
$$+\frac{\Omega}{3}\Sigma\left(\frac{\mathrm{P}_j}{\rho_k}-\frac{\mathrm{P}_k}{\rho_j}\right)\frac{\sin\theta_i}{m_i}+\frac{\Omega^2}{9}\Sigma\frac{1}{m_i}\left(\frac{1}{\rho_j{}^2}+\frac{1}{\rho_k{}^2}-\frac{2\cos\theta_i}{\rho_j\rho_k}\right)$$
$$+\frac{C^2-\Omega^2}{4\delta^2}\Sigma\frac{\rho_i{}^2}{m_i}\sin^2[\omega+\tfrac{1}{3}(\theta_j-\theta_k)]-\Sigma\frac{m_jm_k}{\rho_i}$$

in view of (42), (43), (44), (33), (34) and (53_2). It is understood that in (55_2) one has to express $\theta_1, \theta_2, \theta_3$ and δ by means of (37), (32) and (38) in terms of the three ρ_i. Thus (54_1), (54_2) is a conservative canonical system with 4 degrees of freedom, and with a Hamiltonian function (55_2) which is entirely symmetrical with respect to the three masses.

Remark. If a solution of the system (54_1), (54_2) is known, the corresponding solution of (52) follows from (36_1), (36_2), (36_3) and (30), while the functions

$$\nu=\nu(t),\qquad \mathrm{N}=\mathrm{N}(t),\qquad \iota=\iota(t),\qquad \mathrm{I}=\mathrm{I}(t)$$

follow by a single quadrature. In fact, if ξ_i, η_i, Ξ_i, H_i are known functions of t, then so are $\mathrm{N}\ (\equiv C)$, $\mathrm{I}\ (\equiv 0)$ and ι in view of (51) and (22). Furthermore, on substituting (50) into the first of the equations (49_1) and using then (51), one obtains

$$(56)\qquad \nu'=C\mathrm{M}_{\eta\eta}/\Delta,$$

where $\mathrm{M}_{\eta\eta}$ and Δ are, in view of (17) and (19), known functions of t. Hence $\nu=\nu(t)$ follows from (56) by a quadrature.[10] This corresponds to the fact that the node ν is, in view of (50), an ignorable coördinate in (49_1), (49_2), a fact usually referred to as "the elimination of the node" (Lagrange-Jacobi). The relation which corresponds to (56) in case of the inclination is

$$(57)\qquad \iota'/\sin\iota=C\mathrm{M}_{\xi\eta}/\Delta,$$

a relation obtained in the same way as (56).

V. *Application to the planar problem of three bodies.* It has been assumed so far that the solution is non-planar. Now let the solution be planar

[10] Cf. T. Levi-Civita, *loc. cit.*, [9], pp. 931-933; also Maria Ronchi, "Sulla riduzione esplicita del problema dei tre corpi," *Atti Reale Istituto Veneto*, ser. 9, vol. 1_2 (1917), pp. 1221-1225.

and let (ξ, η) be a barycentric inertial plane containing all three paths. Then, on excluding the case of collinear configurations, (29_1), (29_2), (29_3) are satisfied. Hence the Hamiltonian function is $H = T - U$, where $2T$ is represented by (42), while $U = \Sigma\, m_j m_k \rho_i^{-1}$. It is seen from (22) and (43) that the preservation of the angular momentum is the relation

$$\Omega = C. \tag{58}$$

On substituting (58) into (42) and applying the rule (iii) of Section I to the canonically conjugate pair Ω, ω, where now the coördinate ω does not occur explicitly, it follows that the equations of motion may be written in the form (54_2), where K is given not by (55_2) but by the expression[11]

$$\mathrm{K} = \tfrac{1}{2}\Sigma\, m_i^{-1}(\mathrm{P}_j^2 + \mathrm{P}_k^2 - 2\mathrm{P}_j\mathrm{P}_k \cos\theta_i) - \Sigma\, m_j m_k/\rho_i + \frac{C}{3}\Sigma\left(\frac{\mathrm{P}_j}{\rho_k} - \frac{\mathrm{P}_k}{\rho_j}\right)\frac{\sin\theta_i}{m_i} + \frac{C^2}{9}\Sigma\frac{1}{m_i}\left(\frac{1}{\rho_j^2} + \frac{1}{\rho_k^2} - \frac{2\cos\theta_i}{\rho_j\rho_k}\right). \tag{59}$$

If a solution of this conservative problem (54_2), which has three degrees of freedom, is known, the angle $\omega = \omega(t)$ defined by (36_1) and (30) follows by a quadrature, the reason being that ω is an ignorable coördinate in view of (58).

Remark. Collinear configurations of the three masses have been excluded so far. Such positions (syzygies) occur only for isolated values of t, unless the configuration is collinear for every t. In the latter case the solution is necessarily planar, since[12] the solution is then a homothetic solution, unless the line containing the three masses is independent of t.

THE JOHNS HOPKINS UNIVERSITY.

---

[11] This agrees with the result of Murnaghan, *loc. cit.*, [3], p. 832, since θ_i in (59) and Murnaghan's A_i are exterior and interior angles respectively.

[12] Cf., e. g., Ch. H. Müntz, "Die Aehnlichkeitsbewegungen beim allgemeinen n-Körperproblem," *Mathematische Zeitschrift*, vol. 15 (1923), pp. 169-187.

ON THE VALUES OF THE RIEMANN ζ-FUNCTION ON FIXED LINES $\sigma > 1$.[1]

By Richard Kershner.

For a fixed $\sigma > 1$, let $M(\sigma)$ denote the closure of the set of values taken by the almost periodic function

$$-\log \zeta(s) = \sum_{n=1}^{\infty} \log\,(1 - p_n^{-\sigma} e^{-it\log p_n}), \qquad -\infty < t < +\infty,$$

where $\zeta(s) = \zeta(\sigma + it)$ is the Riemann ζ-function. It has been shown by Bohr [1] that, due to the linear independence of the logarithms of the prime numbers, $M(\sigma)$ is the vectorial sum[2] of the sequence of curves $S(p_1^{-\sigma})$, $S(p_2^{-\sigma}), \cdots$, where $S(r)$ denotes, for a fixed $r < 1$, the curve

$$(1) \qquad w = u + iv = \log\,(1 + re^{i\theta}), \qquad 0 \leqq \theta < 2\pi,$$

in the w-plane. Since every $S(r)$ is easily seen to be a convex Jordan curve, it follows from Bohr's general results (cf. [2]) on the vectorial addition of convex curves that $M(\sigma)$ is,[3] for every fixed $\sigma > 1$, either a closed bounded convex region bordered by a convex Jordan curve $A(\sigma)$ or a closed ring-shaped region bordered by two convex Jordan curves $A(\sigma)$ and $B(\sigma)$, where $B(\sigma)$ is entirely within $A(\sigma)$. Following a suggestion of Wintner, Haviland [5] applied the supporting function (Stützfunktion) of Brunn and Minkowski to the study of $A(\sigma)$, using the fact that the supporting function of $A(\sigma)$ is the sum of the supporting functions of all $S(p_n^{-\sigma})$. A corresponding study of the inner curve $B(\sigma)$, in case $B(\sigma)$ exists, has recently been made by the author [8] and results similar to, but more complicated than, those concerning $A(\sigma)$ were obtained. Actually these investigations treated not only the particular problem of $M(\sigma)$ but the general problem of the vectorial addition of any sequence of convex curves. The object of the present paper is to apply

[1] Received October 28, 1936.

[2] By the vectorial sum $S_1(+)\,S_2(+)\cdots(+)\,S_n$ of n sets S_j of complex numbers is meant the set of numbers w representable in at least one way in the form $w = w_1 + w_2 + \cdots + w_n$ where $w_j \subset S_j$. If $n = \infty$, one means by $S_1(+)S_2(+)\cdots$ the set of points w for which $w = w_1 + w_2 + \cdots$ is a convergent series and $w_j \subset S_j$. In the case under consideration ($\sigma > 1$) this series is absolutely convergent for every choice of $w_j \subset S_j$.

[3] For a short proof of this statement cf. Jessen and Wintner [6], p. 69.

these general results to a detailed discussion of $M(\sigma)$ for varying values of σ. Since the closure $N(\sigma)$ of the values attained by $\zeta(\sigma+it)$ on a fixed line $\sigma>1$ may be obtained from $M(\sigma)$ by a trivial exponential mapping, the results to be obtained may easily be formulated in terms of $\zeta(s)$ instead of $\log\zeta(s)$.

Mention should first be made of the following result of Bohr and Jessen [4]:

(I) *There exists a constant $\sigma_0>1$ such that $M(\sigma)$ is or is not a convex region* (i.e., *$B(\sigma)$ does not or does exist*) *according as $1<\sigma\leqq\sigma_0$ or $\sigma_0<\sigma<+\infty$. This σ_0 is the only root $\sigma>1$ of the equation*

$$\text{(2a)} \qquad \arcsin 2^{-\sigma}=\sum_{n=2}^{\infty}\arcsin p_n^{-\sigma}, \text{ so that } \sigma_0=1.764\cdots.$$

It is easy to see that $S(r)$ has, for every fixed $r<1$, the two lines of symmetry $v=0$ and $u=\frac{1}{2}\log(1-r^2)$, so that $M(\sigma)$ has, for every fixed $\sigma>1$, the two lines of symmetry

$$v=0 \text{ and } u=u_\sigma=\sum_{n=1}^{\infty}\tfrac{1}{2}\log(1-p_n^{-2\sigma}).$$

It has been shown by Kershner and Wintner [7] that $A(\sigma)$ is, for every $\sigma>1$, a regular analytic curve and that the same is true of $B(\sigma)$ up to its possible corners, provided that $B(\sigma)$ exists. In view of the fact (cf. (II) below) that $B(\sigma)$ has at most two corners for any σ, this statement implies a remark of Bohr and Jessen [4] to the effect that $B(\sigma)$ contains no straight segments for any σ.

The following theorem will now be deduced:

(II) *There exists a constant $\sigma^0>\sigma_0$ such that $B(\sigma)$ has exactly two corners or no corners according as $\sigma_0<\sigma<\sigma^0$ or $\sigma^0\leqq\sigma<+\infty$. This σ^0 is the only root $\sigma>1$ of the equation*

$$\text{(2b)} \qquad 2^{-\sigma}=\sum_{n=2}^{\infty}p_n^{-\sigma}, \text{ so that } \sigma^0=1.778\cdots.$$

If $\sigma_0<\sigma<\sigma^0$, both corners of $B(\sigma)$ lie on the real axis.

This result has been announced, without proof, by Bohr and Jessen at the end of their paper [4], where it is stated that the proof of (II) is similar to that of (I). In attempting to verify (II), it was found that (II) may be deduced from the results of the paper [8] as follows:

Let $P(r)$ denote, for a fixed r in $0<r<1$, the circle $|z-1|=r$, so

that $S(r)$ is, by (1), the image of $P(r)$ under the conformal mapping $u + iv = \log z$. Thus, if $z = \rho e^{i\phi}$, $-\frac{1}{2}\pi \leqq \phi \leqq \frac{1}{2}\pi$, then,

$$(3) \qquad u = \log \rho, \quad v = \phi.$$

The ranges of ρ and ϕ, when z describes $P(r)$, are

$$1 - r \leqq \rho \leqq 1 + r \quad \text{and} \quad -\arcsin r \leqq \phi \leqq \arcsin r.$$

Let $\alpha(r, v)$, for $0 < r < 1$, $0 < v \leqq \arcsin r$, denote the angle measured in the positive direction from the u-axis to the normal to $S(r)$ at that point of $S(r)$ cut by the line of constant ordinate v for which u is greater, and let $\rho(r, \alpha)$ be the radius of curvature of $S(r)$ at this same point. Similarly, let $\alpha_\sigma(v)$ be the angle measured in the positive direction from the u-axis to the normal to $B(\sigma)$ at that point of $B(\sigma)$ cut by the line of constant ordinate v for which u is greater, and let $\rho_\sigma(\alpha)$ be the radius of curvature of $B(\sigma)$ at this point, if the radius of curvature is defined there. Clearly, $0 < \alpha_\sigma(v) \leqq \frac{1}{2}\pi$, for the section of $B(\sigma)$ for which $\alpha_\sigma(v)$ has been defined. Due to the above-mentioned symmetry of $B(\sigma)$ it will be sufficient to consider, in the sequel, only that arc of $B(\sigma)$ for which $0 < \alpha_\sigma(v) \leqq \frac{1}{2}\pi$.

It has been shown in [8] that if $\rho_\sigma(\alpha)$ is defined for a given α in $0 < \alpha \leqq \frac{1}{2}\pi$, then, since $\rho(r, \alpha) = \rho(r, \alpha + \pi)$,

$$(4) \qquad \rho_\sigma(\alpha) = \rho(2^{-\sigma}, \alpha) - \sum_{n=2}^{\infty} \rho(p_n^{-\sigma}, \alpha).$$

It was also shown in [8] that, if $\sigma > 1$ is fixed, the radius of curvature $\rho_\sigma(\alpha)$ exists, and so (4) is valid for all positive $\alpha \leqq \frac{1}{2}\pi$ if and only if the expression on the right of (4) is non-negative for all these α and vanishes for isolated values of α, at most. According to Bohr and Jessen, (cf. [3], p. 402),

$$\rho(r, \alpha) = r/\cos v(r, \alpha),$$

where $v(r, \alpha)$ is the unique inverse in $0 < \alpha \leqq \frac{1}{2}\pi$ of $\alpha(r, v)$, so that $0 < v(r, \alpha) \leqq \arcsin r < \frac{1}{2}\pi$. Finally,

$$(5) \qquad \cos v(r, \alpha) = (1 - r^2 \sin^2 \alpha)^{\frac{1}{2}},$$

as seen from (3) and the conformity of the mapping $w = \log z$ (cf. Fig. 1 and Fig. 2 of Bohr and Jessen [4]). Hence, the expression on the right of (4) is

$$(6) \qquad f_\sigma(\alpha) = 2^{-\sigma}(1 - 2^{-2\sigma} \sin^2 \alpha)^{-\frac{1}{2}} - \sum_{n=2}^{\infty} p_n^{-\sigma}(1 - p_n^{-2\sigma} \sin^2 \alpha)^{-\frac{1}{2}},$$

so that, on differentiating with respect to α,

$$f'_\sigma(\alpha) = \sin\alpha\cos\alpha[2^{-3\sigma}(1-2^{-2\sigma}\sin^2\alpha)^{-(3/2)} - \sum_{n=2}^{\infty} p_n^{-3\sigma}(1-2^{-2\sigma}\sin^2\alpha)^{-(3/2)}].$$

Since if $a_n > 0$, then $a_1 \geqq \sum_{n=2}^{\infty} a_n$ implies $a_1^3 \geqq (\sum_{n=2}^{\infty} a_n)^3 > \sum_{n=2}^{\infty} a_n^3$, it follows that $f_\sigma(\alpha) \geqq 0$ implies $f'_\sigma(\alpha) > 0$. Hence, the expression on the right of (4) is a monotone increasing function of α for all those σ for which $f_\sigma(+0) \geqq 0$. Consequently, the radius of curvature $\rho_\sigma(\alpha)$ exists and is represented by (4) for every positive $\alpha \leqq \frac{1}{2}\pi$, if and only if σ is such that the expression on the right of (4) tends to a non-negative limit as $\alpha \to +0$. Now it will be shown that there exists a $\sigma^0 > 1$ such that $f_\sigma(+0)$ is negative, zero or positive according as $1 < \sigma < \sigma^0$, $\sigma = \sigma^0$ or $\sigma^0 < \sigma < +\infty$. First, it is seen from (6) that

$$f_\sigma(+0) = 2^{-\sigma} - \sum_{n=2}^{\infty} p_n^{-\sigma},$$

so that $f_\sigma(+0)$ is positive for large σ, negative for σ near enough to 1 and zero if and only if σ is a root of the equation (2b). Thus, it is sufficient to show that the equation

$$g(\sigma) = 1, \quad \text{where} \quad g(\sigma) = \sum_{n=2}^{\infty} (2/p_n)^\sigma$$

has exactly one root $\sigma = \sigma^0 > 1$. But this is obvious, since $g(\sigma)$ is a continuous, monotone decreasing function for which $g(1+0) = +\infty$, $g(+\infty) = 0$.

In order to complete the proof of (II), it remains to be shown that if $B(\sigma)$ has corners, they lie on the real axis. Suppose, if possible, that $B(\sigma)$ has, for some fixed $\sigma > 1$, a corner not on the real axis. Then it is seen from the double symmetry of $B(\sigma)$ that the α-interval $0 < \alpha \leqq \frac{1}{2}\pi$ contains a subinterval which corresponds to a corner. Let $\alpha_1 \leqq \alpha \leqq \alpha_2$ be this subinterval, so that $0 < \alpha_1 < \alpha_2 \leqq \frac{1}{2}\pi$. It is clear from the considerations of the paper [8], especially from the geometrical construction of $B(\sigma)$ given there, that one can choose this subinterval in such a way that $f_\sigma(\alpha)$ is negative for $\alpha_1 < \alpha < \alpha_2$ and zero for $\alpha = \alpha_1$. Hence $f_\sigma(\alpha)$ is certainly not increasing in the interval $\alpha_1 < \alpha \leqq \frac{1}{2}\pi$. On the other hand, it has been shown above that if $f_\sigma(\alpha_1) = 0$, then $f_\sigma(\alpha)$ is increasing for $\alpha_1 < \alpha \leqq \frac{1}{2}\pi$. This contradiction completes the proof of (II).

Incidentally, the above considerations clearly imply the following result:

(III) *The radius of curvature $\rho_\sigma(\alpha)$ of $B(\sigma)$ is, for every fixed $\sigma > \sigma_0$, defined, continuous and increasing in the α-interval $\alpha_\sigma(+0) < \alpha \leqq \frac{1}{2}\pi$. If $\sigma = \sigma^0$, then $\alpha_\sigma(+0) = 0$ and $\rho_\sigma(+0) = 0$ so that, while the convex curve*

$B(\sigma^0)$ has no corners, its radius of curvature vanishes at the points on the real axis.

If $\sigma \geqq \sigma^0$, then $B(\sigma)$ has no corners by (II), while $B(\sigma)$ does not exist for $\sigma \leqq \sigma_0$ by (I). The following theorem describes the situation in the remaining σ-interval.

(IV) *If $\sigma_0 < \sigma < \sigma^0$ and $\theta = \theta(\sigma)$ denotes the positive angle $(<\pi)$ formed by the two branches of $B(\sigma)$ at a corner of $B(\sigma)$, then $\theta(\sigma)$ is a monotone function of σ and $\theta(\sigma_0 + 0) = 0$, $\theta(\sigma^0 - 0) = \pi$.*

First, it is clear from the definition of $\alpha_\sigma(v)$, as given in the proof of (II), that

$$\theta(\sigma) = \pi - 2\alpha_\sigma(+0). \tag{7}$$

Furthermore, it is clear from the proof of Theorem II_0 in the paper [8] that if Q_a and $P_a{}^n$ denote the points of $B(\sigma)$ and $S(p_n{}^{-\sigma})$ which have the normal inclination α, then Q_α is the vectorial sum of the points $P^1{}_\alpha, P^2{}_{\alpha+\pi}, P^3{}_{\alpha+\pi}, \cdots$. This implies, in view of the definitions of $v_\sigma(\alpha)$ and $v(r,\alpha)$ as given in the proof of (II), that

$$v_\sigma(\alpha) = v(2^{-\sigma}, \alpha) + \sum_{n=2}^{\infty} v(p_n{}^{-\sigma}, \alpha + \pi).$$

Now $v(r,\alpha) = -v(r, \alpha + \pi)$ in view of the summetry of $S(r)$. Consequently, since $\sin v(r,\alpha) = r \sin \alpha$ by (5),

$$v_\sigma(\alpha) = \arcsin\,(2^{-\sigma} \sin \alpha) - \sum_{n=2}^{\infty} \arcsin\,(p_n{}^{-\sigma} \sin \alpha).$$

On the other hand it is clear from the definition of α and from the convexity of $B(\sigma)$ that $\alpha_\sigma(v)$ is increasing in the interval $0 < v \leqq \sum_{n=1}^{\infty} \arcsin p_n{}^{-\sigma}$ so that $\alpha_\sigma(+0)$ is the largest root $(< \tfrac{1}{2}\pi)$ of the equation $v_\sigma(\alpha) = 0$, i. e., of the equation

$$F(\sigma, \alpha) = 0, \tag{8}$$

where

$$F(\sigma,\alpha) = \sum_{n=2}^{\infty} \arcsin\,(p_n{}^{-\sigma} \sin \alpha)/\arcsin\,(2^{-\sigma} \sin \alpha) - 1. \tag{9}$$

First, it will be shown by an argument very similar to that of Bohr and Jessen [4], pp. 42-43, that

$$\partial F(\sigma,\alpha)/\partial\alpha < 0, \qquad 0 < \alpha \leqq \tfrac{1}{2}\pi. \tag{10}$$

In fact, it will be shown that each term

$$\phi_n(\sigma, \alpha) = \arcsin (p_n^{-\sigma} \sin \alpha)/\arcsin (2^{-\sigma} \sin \alpha)$$

of (9) has a negative derivative, $\partial\phi_n/\partial\alpha$. Since the logarithmic derivative $\phi_n^{-1}\partial\phi_n/\partial\alpha$ of ϕ_n is the difference between the logarithmic derivatives of the numerator and the denominator, and since $\phi_n > 0$, in order to show that $\partial\phi_n/\partial\alpha < 0$ it is sufficient to prove that the logarithmic derivative $\psi^{-1}\partial\psi/\partial\alpha$ of the function $\psi(\alpha, \lambda) = \arcsin (\lambda \sin \alpha)$ is, for each fixed α in the interval $0 < \alpha \leqq \frac{1}{2}\pi$, an increasing function of λ in $0 < \lambda < 1$. Now the latter logarithmic derivative is

$$\text{(11)} \qquad \frac{\lambda \cos \alpha}{\sqrt{1 - \lambda^2 \sin^2 \alpha} \arcsin (\lambda \sin \alpha)}, \quad \text{i.e.,} \quad \cot \alpha \frac{\tan y}{y},$$

where $y = \arcsin (\lambda \sin \alpha)$, so that y is an increasing function of λ. Since $(\tan y)/y$ is an increasing function of y in $0 < y \leqq \frac{1}{2}\pi$ it follows that the logarithmic derivative (11) is an increasing function of λ. This completes the proof of (10).

Now it has been pointed out above that $\alpha = \alpha_\sigma(+0)$ is a root of the equation (8). On the other hand, $\alpha = \alpha_\sigma(+0)$ lies in the α-range admitted in (10), since $\alpha_\sigma(+0) = 0$ would imply that $B(\sigma)$ has no corners, whereas $\sigma_0 < \sigma < \sigma^0$ by assumption. On substituting $\alpha = \alpha_\sigma(+0)$ into (8), it follows, from (10) and from the obvious existence of continuous partial derivatives of the function (9), that the function $\alpha_\sigma(+0)$ of σ possesses the continuous derivative

$$\text{(12)} \quad d\alpha_\sigma(+0)/d\sigma = -\partial F(\sigma, \alpha)/\partial\sigma : \partial F(\sigma, \alpha)/\partial\alpha \quad (\alpha = \alpha_\sigma(+0)).$$

Now it is easily seen from the considerations of Bohr and Jessen ([4], pp. 42-43) that

$$\text{(13)} \qquad \partial F(\sigma, \alpha)/\partial\sigma < 0 \qquad (0 < \alpha \leqq \tfrac{1}{2}\pi,\ \sigma_0 < \sigma < \sigma^0).$$

On comparing (12) with (10) and (13) it follows that $d\alpha_\sigma(+0)/d\sigma < 0$, so that $\theta(\sigma)$ is, in view of (7) an increasing function of σ in the interval $\sigma_0 < \sigma < \sigma^0$. In order to complete the proof of (IV) it is sufficient to notice that $\theta(\sigma_0 + 0) = 0$ is obvious, if one compares the equation (2a) which defines σ_0 with the relations (8), (9), while $\theta(\sigma^0 - 0) = \pi$ similarly follows from the equation (2b) which defines σ^0, if one lets $\alpha \to 0$ in (9).

With regard to the shape of the region $M(\sigma)$ for large values of σ one may say the following:

(V) *The region $M(\sigma)$ is, for large values of σ, asymptotically a circle of radius $2^{-\sigma}$.*

In fact, let $h(r, \alpha)$ be the supporting function of $S(r)$ so that $h(r, \alpha) = u \cos \alpha + v \sin \alpha$, where (u, v) is the point of $S(r)$ for which the normal inclination has the value α. Then on introducing for u and v their expressions in terms of r and α and expanding in powers of r, it may easily be shown that

$$h(r, \alpha) \sim r \text{ as } r \to 0,$$

uniformly in α. Now if $g(\sigma, \alpha)$ is the supporting function of $A(\sigma)$, then, according to the rule of Haviland ([5], p. 333), $g(\sigma, \alpha) = \sum_{n=1}^{\infty} h(p_n^{-\sigma}, \alpha)$, so that

$$(14) \qquad g(\sigma, \alpha) \sim 2^{-\sigma} + \sum_{n=2}^{\infty} p_n^{-\sigma} \sim 2^{-\sigma}; \qquad \sigma \to +\infty;$$

uniformly for all α.

Similarly, if $k(\sigma, \alpha)$ is the supporting function of $B(\sigma)$ and if $\sigma > \sigma^0$, then, by [8], p. 743, $k(\sigma, \alpha) = h(2^{-\sigma}, \alpha) - h(p_n^{-\sigma}, \alpha + \pi)$, so that

$$(15) \qquad k(\sigma, \alpha) \sim 2^{-\sigma} - \sum_{n=2}^{\infty} p_n^{-\sigma} \sim 2^{-\sigma}, \qquad \sigma \to +\infty;$$

uniformly for all α. Since (14) and (15) hold uniformly for all α, the result (V) follows.

In conclusion, the following obvious remark concerning the shape of $B(\sigma)$, as $\sigma \to \sigma_0$, may be made.

(VI) *The ratio $R(\sigma)$ of the diameter of $B(\sigma)$ along the u-axis to the diameter parallel to the v-axis tends to infinity as $\sigma \to \sigma_0$.*

In fact, a trivial application of the mean value theorem gives $R(\sigma) > \tan \alpha_\sigma(+0)$, while $\tan \alpha_\sigma(+0) \to +\infty$ as $\sigma \to \sigma_0$, by (IV) and (7).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] H. Bohr, "Sur la fonction $\zeta(s)$ dans le demi-plan $\sigma > 1$," *Comptes Rendus*, vol. 154 (1912), pp. 1078-1081.

[2] H. Bohr, "Om Addition af uendelig mange konvekse Kurver," *Danske Videnskabernes Selskab, Forhandlinger*, 1913, pp. 325-366.

[3] H. Bohr and B. Jessen, "Om Sandsynlighedsfordelinger ved Addition af konvekse Kurver," *Danske Videnskabernes Selskab*, Skrifter, (8), vol. 12, no. 3 (1929).

[4] H. Bohr and B. Jessen, "On the distribution of the values of the Riemann zeta function," *American Journal of Mathematics*, vol. 58 (1936), pp. 35-45.

[5] E. K. Haviland, "On the addition of convex curves in Bohr's theory of Dirichlet series," *American Journal of Mathematics*, vol. 55 (1933), pp. 332-334.

[6] B. Jessen and A. Wintner, "Distribution functions and the Riemann zeta function," *Transactions of the American Mathematical Society*, vol. 38 (1935), pp. 48-88.

[7] R. Kershner and A. Wintner, "On the boundary of the range of values of $\zeta(s)$," *American Journal of Mathematics*, vol. 58 (1936), pp. 421-426.

[8] R. Kershner, "On the addition of convex curves," *American Journal of Mathematics*, vol. 58 (1936), pp. 737-747. Formula (ii), p. 745, should read

$$\sqrt{A(C_I)} \leqq \sqrt{A(C_1)} - \sum_{j=2}^{n} \sqrt{A(C_j)}.$$

CONVOLUTIONS OF DISTRIBUTIONS ON CONVEX CURVES AND THE RIEMANN ZETA FUNCTION.*

By E. R. van Kampen and Aurel Wintner.

Introduction. It is known that the almost periodic function

$$x + iy = z = z(t) = \log \zeta(\sigma + it), \qquad -\infty < t < +\infty, \tag{1}$$

where $\sigma > 1$ is fixed, has an absolutely continuous asymptotic distribution function the density of which possesses continuous partial derivatives of arbitrarily high order.[1] The same holds for this density $\delta = \delta^\sigma = \delta^\sigma(x, y)$ in the case $\sigma = 1$, while if $\frac{1}{2} < \sigma < 1$, then $\delta^\sigma(x, y)$ is a transcendental entire function of two variables.[2] These results are proved by estimating the order of magnitude of the Fourier transform in the infinity.[3] It is not known whether or not $\delta^\sigma(x, y)$ is, in the case $\sigma = 1$, regular analytic in the whole real (x, y)-plane, or at least in some portions of it. In what follows, it will be assumed that $\sigma > 1$. In this case the spectrum of the asymptotic distribution function of (1) is a bounded set, so that $\delta^\sigma(x, y)$, being identically zero without the spectrum, vanishes with all its derivatives on the boundary of the spectrum, and so $\delta^\sigma(x, y)$ cannot be regular analytic there. It may be mentioned that the boundary of the spectrum consists of regular analytic curves.[4] The question of subregions of analyticity within the spectrum cannot be discussed by the method of Fourier transforms mentioned above. In fact, this method yields the regular analyticity of $\delta^\sigma(x, y)$ either for every (x, y) or for no (x, y).

In what follows, there will be delimited within the spectrum subregions of regular analyticity for the density $\delta^\sigma = \delta^\sigma(x, y)$. It turns out that there exists a sequence $\alpha_1, \alpha_2, \cdots$ of numbers $\alpha_k > 1$ such that $\delta^\sigma(x, y)$ is for every $\sigma > \alpha_k$ regular analytic in a certain number, say N_k, of mutually disjoint ring-shaped subregions of the spectrum, and that $N_k \to +\infty$ as $k \to +\infty$. Since the N_k rings form a disconnected set, it cannot be stated that the corresponding N_k regular functions of two variables x, y are analytic continuations of each

* Received November 16, 1936.

[1] Wintner [16], pp. 328-329; Jessen and Wintner [10], Theorems 19 and 30; Haviland and Wintner [9].

[2] Jessen and Wintner [10], Theorems 19 and 30.

[3] Wintner [17]; Jessen and Wintner [10], p. 54.

[4] Kershner and Wintner [15].

other. There are some indications to the effect that the Jordan curves which form the boundary of the ring-shaped regions to be determined consist of singular points of $\delta^\sigma(x, y)$; but we cannot prove this. It is not stated that the N_k rings to be determined contain all subregions of analyticity within the spectrum.

The results in the Appendix concern the existence and a discussion of singularities within the spectrum of the asymptotic distribution functions of real almost periodic functions with linearly independent frequencies.

The proof to be given for the existence of rings of analyticity of $\delta^\sigma(x, y)$ for $\sigma > \alpha_k$ depends largely on the geometry of convolutions of distributions on convex curves, so that the essential part of the paper consists of necessary refinements of certain elementary geometrical results contained in the literature.[5] In fact, the results to be obtained on the geometry of the convolutions in question will make it possible to establish the existence of rings of analyticity by means of an argument previously applied [6] in the case of almost periodic functions with linearly independent frequencies, i. e., in the case of infinite convolutions of circular equidistributions. It may be mentioned that for the convergence of the infinite convolutions in question there exists, in case of a bounded spectrum, an entirely geometrical criterion, involving not the distributions along the partial convex curves but merely the latter curves themselves (Section 4).

On applying a trivial exponential mapping, the results obtained for the asymptotic distribution function of (1) yield corresponding results for the asymptotic distribution function of

$$x + iy = z = z(t) = \zeta(\sigma + it), \qquad -\infty < t < +\infty.$$

The method used in the case of $\log \zeta(s)$ will be applied to $\zeta'(s)/\zeta(s)$ also. While in the latter case the geometry of the problem [7] is as simple as in the case of circular equidistributions, the problem of the densities is more involved.

1. *The convolution of two distributions on convex curves.* Let Θ_1, Θ_2 be two oriented circles and let θ_j denote the angular coördinate on Θ_j. If A_1, A_2 are subsets of Θ_1, Θ_2 respectively, let $A_1 \times A_2$ denote the product space of A_1 and A_2, so that, in particular, $\Theta_1 \times \Theta_2$ denotes a torus on which a point is characterized by a pair (θ_1, θ_2) of angular coördinates. If A, A_j, where $j = 1, 2$, are Borel sets on $\Theta_1 \times \Theta_2$, Θ_j respectively, put

[5] Bohr [2], [3]; Bohr and Jessen [5]; Haviland [8]; Bohr and Jessen [6]; Kershner [11], [12].

[6] Kershner and Wintner [14].

[7] Bohr [1], Burrau [7].

$$(2) \qquad \mu(\mathrm{A}) = \frac{\text{meas A}}{\text{meas } \Theta_1 \times \Theta_2}, \qquad \lambda(\mathrm{A}_j) = \frac{\text{meas A}_j}{\text{meas } \Theta_j}, \qquad (j = 1, 2),$$

where the measures refer to those on $\Theta_1 \times \Theta_2$, Θ_j respectively, when $\Theta_1 \times \Theta_2$, Θ_j are thought of as represented by means of the parameters $(\theta_1, \theta_2), \theta_j$. Thus $\mu(\mathrm{A})$, $\lambda(\mathrm{A}_j)$ are non-negative absolutely additive set-functions which are defined for all Borel sets A, A_j on $\Theta_1 \times \Theta_2$, Θ_j respectively and have the total variation 1. The distribution functions $\mu(\mathrm{A})$, $\lambda(\mathrm{A}_j)$ represent equidistributions with constant densities $(\frac{1}{2\pi})^2$, $\frac{1}{2\pi}$ on $\Theta_1 \times \Theta_2$, Θ_j respectively. It is clear that if A_1, A_2 are Borel sets on Θ_1, Θ_2 respectively, then the product space $\mathrm{A}_1 \times \mathrm{A}_2$ is a Borel set on the torus $\Theta_1 \times \Theta_2$ and

$$\mu(\mathrm{A}_1 \times \mathrm{A}_2) = \lambda(\mathrm{A}_1)\lambda(\mathrm{A}_2).$$

Let a function

$$(3) \qquad x + iy = z = z_j(\theta_j) \equiv \xi_j(\theta_j) + i\eta_j(\theta_j)$$

defined on Θ_j, where $j = 1, 2$, be an admissible parameter representation of a convex curve S_j in the z-plane, the adjective "admissible" being meant in the usual sense. In other words, on denoting by T_j the transformation (3) of the circle Θ_j into the convex curve S_j, it is assumed that T_j is a topological transformation and that the function (3) of θ_j has a continuous derivative which is nowhere zero:

$$(4) \qquad |z'_j(\theta_j)|^2 = \xi'_j(\theta_j)^2 + \eta'_j(\theta_j)^2 \neq 0.$$

It may be assumed that the orientation of Θ_j, when transplanted by means of T_j onto S_j, determines on S_j that orientation which is positive in the z-plane, so that the positively oriented normal at a point of S_j is the exterior normal at that point.

In addition to the assumption that (3) is an admissible parameter representation of the convex curve S_j, let it be supposed that there exists for every $\omega = \arg z$ exactly one point θ_j on Θ_j, say

$$(5) \qquad \theta_j = \vartheta_j(\omega), \qquad (j = 1, 2),$$

such that the positive normal at the point $z = z_j(\theta_j)$ of S_j has the normal inclination ω. In other words, it will be supposed that the convex curve S_j does not contain rectilinear segments. It is clear that (5) is an orientation-preserving topological transformation of the unit circle $z = e^{i\omega}$ of the z-plane into the circle Θ_j.

For a given Borel set E in the z-plane, let $\phi_j(E)$ denote $\lambda(T_j^{-1}(ES_j))$,

12

i. e., the value of the set-function $\lambda(\mathrm{A}_j)$ in (2) for the Borel set $\mathrm{A}_j = T_j^{-1}(ES_j)$ into which the inverse of the transformation (3) transforms the common part of the Borel set E and of the convex curve S_j. Thus $\phi_j(E)$ is a continuous, but not absolutely continuous, distribution function which has S_j as spectrum and is defined for all Borel sets E in the z-plane. The convolution [8] of the distribution functions ϕ_1, ϕ_2 will be denoted by ψ_2:

$$(6) \qquad \psi_2(E) = \psi_2 = \phi_1 * \phi_2.$$

For a given pair of functions (3), let T denote the transformation

$$(7) \qquad x + iy = z(\theta_1, \theta_2) \equiv \sum_{j=1}^{2} z_j(\theta_j) \equiv \sum_{j=1}^{2} \xi_j(\theta_j) + i \sum_{j=1}^{2} \eta_j(\theta_j)$$

of the torus $\Theta_1 \times \Theta_2$ into a subset of the z-plane. This subset, $T(\Theta_1 \times \Theta_2)$, of the z-plane is what is called the vectorial sum of the convex curves S_1, S_2 and will be denoted, in contradistinction to the logical sum $S_1 + S_2$, by $S_1(+)S_2$. The Jacobian of the transformation T is, according to (7),

$$(8) \qquad J = J(\theta_1, \theta_2) = \frac{\partial(x, y)}{\partial(\theta_1, \theta_2)} = \begin{vmatrix} x'_1(\theta_1) & x'_2(\theta_2) \\ y'_1(\theta_1) & y'_2(\theta_2) \end{vmatrix}.$$

Now [9] the convolution (6) is in the same way the transplantation of the equidistribution $\mu(A)$ on $\Theta_1 \times \Theta_2$ onto the z-plane under the mapping T as the distribution function $\phi_j = \phi_j(E)$ is the transplantation of the equidistribution $\lambda(\mathrm{A}_j)$ on Θ_j onto the z-plane under the mapping T_j, where $j = 1, 2$. More explicitly, $\psi_2(E)$ is, for every Borel set E in the z-plane, equal to $\mu(T^{-1}(E))$, where the Borel set $\mathrm{A} = T^{-1}(E)$ is meant to be the inverse image of E under the transformation T, i. e., the set of those points (θ_1, θ_2) of $\Theta_1 \times \Theta_2$ which are transformed by the continuous, but not topological, transformation (7) into points z of E. In particular, $S_1(+)S_2$ is the spectrum of the distribution function (6). It is also seen that if

$$\bar{\Lambda} = \Theta_1 \times \Theta_2 - \sum_{k=1}^{m} \Lambda_k,$$

where $\Lambda_1, \cdots, \Lambda_m$ are mutually disjoint Borel sets on $\Theta_1 \times \Theta_2$, then

$$(9a) \qquad \psi_2(E) = \psi_2(E; \bar{\Lambda}) + \sum_{k=1}^{m} \psi_2(E; \Lambda_k),$$

where

$$(9b) \qquad \psi_2(E; \Lambda) = \mu(\mathrm{A}\Lambda), \qquad \mathrm{A} = T^{-1}(E),$$

[8] As to terminology and notations, cf. Jessen and Wintner [10], Section 2.

[9] Jessen and Wintner [10], p. 84.

so that $\psi_2(E;\Lambda)$ denotes the contribution of a portion Λ of $\Theta_1 \times \Theta_2$ to $\psi_2(E) = \psi_2(E;\Theta_1 \times \Theta_2)$.

2. *The set Ω and its complement on the torus.* Let Ω denote the set of those points (θ_1, θ_2) of the torus $\Theta_1 \times \Theta_2$ at which the Jacobian (8) of the transformation T vanishes. Thus the continuous transformation T defined by (7) is locally topological on the set

$$\bar{\Omega} = \Theta_1 \times \Theta_2 - \Omega,$$

which, Ω being obviously closed, is an open subset of $\Theta_1 \times \Theta_2$.

In order to describe Ω in terms of the two periodic functions (5), let there be defined on the torus $\Theta_1 \times \Theta_2$ two disjoint rectifiable Jordan curves Ω^+, Ω^- by means of the parameter representations

$$\text{(10a)} \qquad \Omega^+ : \theta_1 = \vartheta_1(\omega), \quad \theta_2 = \vartheta_2(\omega);$$

$$\text{(10b)} \qquad \Omega^- : \theta_1 = \vartheta_1(\omega), \quad \theta_2 = \vartheta_2(\omega) + \pi.$$

Then Ω is the logical sum

$$\text{(11)} \qquad \Omega = \Omega^+ + \Omega^-.$$

In fact, it is clear from (4) and (8) that $J(\theta_1, \theta_2)$ vanishes at the point (θ_1, θ_2) of $\Theta_1 \times \Theta_2$ if and only if the oriented normal to S_1 at the T_1-image of the point θ_1 of Θ_1 is either parallel or anti-parallel to the oriented normal to S_2 at the T_2-image of the point θ_2 of Θ_2. This means in view of the definition of the functions (5) that $J(\theta_1, \theta_2) = 0$, i.e. $(\theta_1, \theta_2) \subset \Omega$, holds if and only if (θ_1, θ_2) is a point either of the curve (10a) or of the curve (10b). That the components Ω^+, Ω^- of the set (11) are disjoint rectifiable Jordan curves, is seen from (10a), (10b) and from the fact that each of the functions (5) defines, by Section 1, a topological transformation of a circle into another circle.

Let R denote the portion of $S_1 (+) S_2$ not contained in the T-image of (11):

$$\text{(12)} \qquad R = [S_1 \, (+) \, S_2] - T(\Omega).$$

It is easy to see that R is an open set. In fact, $\bar{\Omega}$ is an open set on the torus $\Theta_1 \times \Theta_2$ and T is locally topological on $\bar{\Omega}$, so that $T(\bar{\Omega})$ is an open set in the z-plane. On the other hand, Ω is a closed subset of the compact set $\Theta_1 \times \Theta_2$ on which T is continuous, so that $T(\Omega)$ is a closed set in the z-plane. Now

$$S_1 \, (+) \, S_2 = T(\Theta_1 \times \Theta_2) = T(\Omega + \bar{\Omega}) = T(\Omega) + T(\bar{\Omega}),$$

so that R is, in view of (12), the set of those points z which are in $T(\bar{\Omega})$ but not in $T(\Omega)$. Since $T(\bar{\Omega})$ is open and $T(\Omega)$ is closed, it follows that R is open.

Next, if $z = z_0$ is any fixed point of R and $\epsilon = \epsilon(z_0) > 0$ is so small that the circle

$$U : |z - z_0| \leqq \epsilon \tag{13a}$$

is contained in R, then there exists on the torus $\Theta_1 \times \Theta_2$ a finite number, say $m = m(U)$, of mutually disjoint closed sets

$$\Lambda_1, \cdots, \Lambda_k, \cdots, \Lambda_m \tag{13b}$$

such that every Λ_k is transformed by T into U in a topological way, while no point of

$$\bar{\Lambda} = \Theta_1 \times \Theta_2 - \sum_{k=1}^{m} \Lambda_k \tag{13c}$$

is transformed by T into a point of U. In fact, since T is locally topological on $\bar{\Omega}$ and R was seen to be a subset of $T(\bar{\Omega})$, it is clear from the monodromy theorem that it is sufficient to prove that the T^{-1}-image of any fixed point z_0 of R is a finite subset of $\bar{\Omega}$. Now suppose, if possible, that the T^{-1}-image of a z_0 is not a finite set and has, therefore, a cluster point ζ on the torus $\Theta_1 \times \Theta_2$. Since T is locally topological on $\bar{\Omega}$, the cluster point ζ cannot be a point of $\bar{\Omega}$, so that ζ must be a point of the complement Ω of $\bar{\Omega}$. This is, however, impossible, since the T^{-1}-image of the point z_0 is a closed set contained in $\bar{\Omega}$.

It has been shown by Bohr [10] that the vectorial sum S_1 (+) S_2 mentioned in Section 1, i. e. the T-image of the torus $\Theta_1 \times \Theta_2$, is either a closed bounded convex region or a closed bounded ring-shaped region bordered by two convex curves which have no point in common. In the first case let C_2 denote the convex curve which forms the boundary of S_1 (+) S_2; in the second case let C_2 and D_2 denote the convex curves which form the outer and inner boundary of S_1 (+) S_2 respectively. Now the convex curve C_2 (in both cases) and the convex curve D_2 (in the second case) are related to the curves Ω^+ and Ω^- on the torus $\Theta_1 \times \Theta_2$ as follows:

$$T(\Omega^+) = C_2, \qquad T^{-1}(C_2) = \Omega^+,$$

and, if D_2 exists,

$$T(\Omega^-) \supset D_2, \qquad T^{-1}(D_2) \subset \Omega^-,$$

where

$$T(\Omega^-) = D_2, \qquad T^{-1}(D_2) = \Omega^-$$

[10] Bohr [2]; cf. also the presentation in Jessen and Wintner [10], Section 9.

holds if and only if D_2 is free of corners. These facts may be obtained by comparing the above considerations with those of Haviland [11] and Kershner [12] with regard to C_2 and D_2 respectively. Hence if D_2 exists and is free of corners, then the open set R defined by (12) and (11) is identical with the open ring-shaped region bordered by the boundary curves C_2, D_2 of the vectorial sum $S_1 \, (+) \, S_2$.

3. *The density of the convolution* (6). In this Section 3 there will be obtained for the distribution function ψ_2 defined by (6) an explicit representation in terms of the Jacobian (8). This representation will not only imply that ψ_2 is, in contrast with ϕ_1 and ϕ_2, an absolutely continuous distribution function but will also show that the density $\delta_2 = \delta_2(x, y)$ of ψ_2 is always continuous on the open set R, finally that $\delta_2(x, y)$ is regular analytic on R in case the functions (3) are regular analytic on Θ_1, Θ_2. Incidentally, $\delta_2(x, y)$ will be seen to be non-bounded on any open set containing a boundary point of R. A proof of the absolute continuity of ψ_2 has been indicated by Bohr.[13] The sharper results to be obtained, which are indispensable for the final purpose of the present paper, will be proved by considerations suggested by Bohr's remarks but will be more elementary in nature, since no use need be made of Lebesgue integrals.

For any fixed point z_0 of the open set R, choose an $\epsilon > 0$ so small that the circle U defined by (13a) is contained in R. Then the T^{-1}-image of U consists, by Section 2, of the $m = m(U)$ closed sets (13b) each of which is transformed by T into U in a topological way. Let

$$(14) \qquad \theta_1 = \alpha_k(x, y), \qquad \theta_2 = \beta_k(x, y), \qquad (k = 1, 2, \cdots, m),$$

denote the topological transformation T^{-1} of U into Λ_k. Since every Λ_k is a closed subset of $\bar{\Omega}$, it is clear from the definition of Ω given at the beginning of Section 2 that the reciprocal value of the Jacobian (8) is bounded on each of the m sets Λ_k. Finally, as pointed out in Section 1, the convolution (6) may be obtained by transplanting the distribution function $\mu(\mathrm{A})$ from the torus $\Theta_1 \times \Theta_2$ onto the z-plane by means of the transformation T; and $\mu(\mathrm{A})$ is the equidistribution on $\Theta_1 \times \Theta_2$, having the constant density $(\frac{1}{2\pi})^2$.

On comparing these facts with the definition (9b) of the symbol $\psi_2(E; \Lambda)$, it is clear from (14), (8) and from the transformation formula of Riemann double integrals that, for any rectangle E contained in the circle (13a),

[11] Haviland [8].
[12] Kershner [11].
[13] Bohr [3].

$$\psi_2(E;\Lambda_k)=\iint\limits_E |\,4\pi^2 J(\alpha_k(x,y),\beta_k(x,y))\,|^{-1}\,dx\,dy, \tag{15a}$$

where $k=1,\cdots,m$. Similarly, since the T^{-1}-image of U consists of the closed sets (13a) and has, therefore, neither a point nor a cluster point in the set (13c), the application of the definition (9b) to the set $\Lambda=\bar{\Lambda}$ clearly gives

$$\psi_2(E;\bar{\Lambda})=0 \tag{15b}$$

for any rectangle E contained in the circle (13a). On substituting (15a), (15b) into (9a), it is seen that

$$\psi_2(E)=\iint\limits_E \sum_{k=1}^{m} |\,4\pi^2 J(\alpha_k(x,y),\beta_k(x,y))\,|^{-1}\,dx\,dy, \tag{16}$$

if E is a rectangle contained in the circle (13a). Since z_0 in (13a) is any point of R, it follows by writing z instead of z_0 that the set-function $\psi_2(E)$ is absolutely continuous on the subset R of $S_1\,(+)\,S_2$ and has at any point $z=x+iy$ of R the density

$$\delta_2(x,y)=\sum_{k=1}^{m} |\,4\pi^2 J(\alpha_k(x,y),\beta_k(x,y))\,|^{-1}, \tag{17}$$

where the integer m or the functions (14) may vary if and only if one chooses the point $z=x+iy$ on distinct connected parts of the open set R. On the other hand, $\psi_2(E)$ is absolutely continuous on the complement of the spectrum $S_1\,(+)\,S_2$ with respect to the whole z-plane, the density $\delta_2(x,y)$ being identically zero on this complement. Hence it is clear from (12) that $\psi_2(E)$ is an absolutely continuous distribution function if and only if

$$\psi_2(T(\Omega))=0. \tag{18}$$

Now, while it is clear from (17) and from the definition of Ω that the density $\delta_2(x,y)$ is non-bounded in any vicinity of any point of the boundary $T(\Omega)$ of R, it is easy to see that (18) is true.

First, since $\psi_2(E)$ is the transplantation of the equidistribution $\mu(\mathrm{A})$ on $\Theta_1\times\Theta_2$ onto the z-plane, (18) is clearly equivalent to

$$\mu(\mathrm{A}_0)=0,\ \text{where } \mathrm{A}_0=T^{-1}(T(\Omega)). \tag{19}$$

Needless to say, one cannot write Ω instead of $T^{-1}(T(\Omega))$, since T^{-1} is multivalued on the T-image of Ω. Now Ω consists, according to (11), of two rectifiable Jordan curves on the torus $\Theta_1\times\Theta_2$, so that, if $\epsilon>0$ is given, one

can choose on $\Theta_1 \times \Theta_2$ an open set A_ϵ which contains Ω, consists of two ring-shaped regions bounded by rectifiable Jordan curves and is such that $\mu(A_\epsilon) < \epsilon$. Hence, in order to prove (19), it is sufficient to show that the common part $A_0\bar{A}_\epsilon$ of $A_0 = T^{-1}(T(\Omega))$ and of the complement $\bar{A}_\epsilon = \Theta_1 \times \Theta_2 - A_\epsilon$ of A_ϵ has a vanishing Jordan content on $\Theta_1 \times \Theta_2$. Since the reciprocal value of the Jacobian (8) is bounded on $\bar{A}_\epsilon$, it is thus sufficient to show that $T(\Omega)$ has a vanishing planar Jordan content. But this is obvious, since the Jacobian (8) is bounded on $\Theta_1 \times \Theta_2$ and Ω consists of two rectifiable Jordan curves on $\Theta_1 \times \Theta_2$.

The result thus proved may be formulated as follows: On defining in the z-plane a function $\delta_2(x, y) \geqq 0$ by placing it equal to the expression (17) or to 0 according as $z = x + iy$ is or is not in R, the convolution (6) may be represented for any Borel set E in the form

$$\psi_2(E) = \iint\limits_E \delta_2(x, y)\, dx\, dy. \tag{20}$$

Furthermore, if E is a closed rectangle, then (20) exists as a proper or as an improper Riemann integral according as E does or does not contain a boundary point of R. Incidentally, the above considerations imply that the closure of the set (12) is the spectrum S_1 (+) S_2 of (6).

Remark. Suppose that the curves (3) are defined and satisfy the requirements of Section 1 not only for $j = 1, 2$ but for $j = 1, 2, \cdots, n$. Then if $\phi_1, \phi_2, \cdots, \phi_n$ denote corresponding distribution functions, the convolution $\phi_1 * \phi_2 * \cdots * \phi_n$ is absolutely continuous for every $n \geqq 2$ and has a continuous density for every $n \geqq 4$. In fact, the first statement is true for $n = 2$, and so it is true for every $n > 2$, since the convolution of an absolutely continuous and of an arbitrary distribution function is absolutely continuous.[14] For similar reasons, it is sufficient to show that the second statement, which has been proved by Bohr and Jessen [15] under somewhat more restrictive conditions and on using elementary geometrical considerations, is true in the lowest case, $n = 4$. Now if $\delta_2{}^{\mathrm{I}}$ and $\delta_2{}^{\mathrm{II}}$ denote the densities of $\phi_1 * \phi_2$ and $\phi_3 * \phi_4$ respectively, $\phi_1 * \phi_2 * \phi_3 * \phi_4$ is absolutely continuous with the density

$$D(x, y) = \int_{-\infty}^{+\infty}\!\!\int \delta_2{}^{\mathrm{I}}(x - \xi, y - \eta)\, \delta_2{}^{\mathrm{II}}(\xi, \eta)\, d\xi\, d\eta,$$

[14] Cf., e. g., Kershner and Wintner [13], p. 543, footnote.

[15] Bohr and Jessen [5].

so that, since $\delta_2^{\mathrm{I}}(\xi,\eta)$, $\delta_2^{\mathrm{II}}(\xi,\eta)$ are non-negative and have finite integrals $(=1)$ over the whole (ξ,η)-plane, $D(x,y)$ is everywhere continuous.

4. *Infinite convolutions.* Let $S_1, S_2, \cdots$ be an infinite sequence of convex curves S_m in the z-plane such that every S_m has a continuously turning tangent and does not contain rectilinear segments. Let

$$(21)\qquad S_m : x + iy = z = z_m(\theta) \equiv \xi_m(\theta) + i\eta_m(\theta); \qquad 0 \leqq \theta < 2\pi,$$

be an admisssible parameter representation of S_m in terms of an angular parameter θ and let $\phi_m = \phi_m(E)$ denote the corresponding distribution function defined for all Borel sets E of the z-plane (cf. Section 1). Thus S_m is the spectrum of ϕ_m, and so the spectrum of the convolution

$$(22)\qquad \psi_n(E) = \psi_n = \psi_1 * \cdots * \psi_n$$

is the vectorial sum

$$(23)\qquad V_n = S_1\,(+)\cdots(+)\,S_n,$$

where $n = 2, 3, \cdots$.

Let d_m denote the diameter of S_m, the diameter of a set being meant in the usual sense. Suppose that

$$(24)\qquad \sum_{m=1}^{\infty} d_m \text{ is convergent.}^{16}$$

Then the infinite convolution

$$(25)\qquad \psi = \phi_1 * \phi_2 * \cdots$$

is convergent if and only if there exists on or within every S_m a point z_m such that

$$(26)\qquad \sum_{m=1}^{\infty} z_m$$

is convergent, in which case (26) is uniformly convergent when every z_m varies on or within S_m. This clearly implies that if (24) is satisfied, then the infinite convolution (25) is absolutely convergent if and only if (26) is absolutely convergent for at least one choice of $z_1, z_2, \cdots$ on or within $S_1, S_2, \cdots$, in which case (26) is absolutely-uniformly convergent for all choices of $z_1, z_2, \cdots$ on or within $S_1, S_2, \cdots$. In particular, the infinite convolution (25) is absolutely convergent whenever (24) is satisfied and every S_m surrounds

[16] While the direction of a maximal cord of S_m may vary with m, an elementary consideration shows that (24) is satisfied if and only if $v_n <$ const., where v_n denotes the diameter of the set (23).

the origin of the z-plane. The interest of these criteria lies in the fact that no mention is made of the parameter representations (21) which determine the distribution functions ϕ_m occurring in (25), i. e., that if (24) is satisfied, then the convergence or absolute convergence of (25) depends only on the curves $S_1, S_2, \cdots$ themselves.

First, on placing

$$(27) \qquad \gamma_m = \frac{1}{2\pi} \int_0^{2\pi} z_m(\theta)\, d\theta$$

and noting that (21) is a parameter representation of the convex curve S_m, it is clear that, d_m being the diameter of S_m, every point $z = z_m$ on or within S_m satisfies the inequality

$$(28) \qquad | z_m - \gamma_m | \leqq d_m.$$

Since the point $z = z_m(\theta)$ is on S_m, it follows that $d_m{}^2$ is an upper bound for the value of

$$(29) \qquad \mu_m = \frac{1}{2\pi} \int_0^{2\pi} | z_m(\theta) - \gamma_m |^2\, d\theta.$$

Hence the assumption (24) implies that

$$(30) \qquad \sum_{m=1}^{\infty} \mu_m \text{ is convergent.}$$

It is also seen from (28) and (24) that (26) is uniformly convergent for all choices of $z_1, z_2, \cdots$ on or within $S_1, S_2, \cdots$ whenever (26) is convergent for a single choice of $z_1, z_2, \cdots$ on or within $S_1, S_2, \cdots$; for instance, whenever

$$(31) \qquad \sum_{m=1}^{\infty} \gamma_m$$

is a convergent series (in fact, $z = \gamma_m$ is, by (27), a point within S_m). Now if (31) is convergent, then it is clear from (21), (28) and (24) that all S_m are contained in a sufficiently large circular disk, and so (30) and the convergence of (31) imply[17] the convergence of (25). Conversely, if (25) is convergent, then[18] the convex curve S_m tends, as $m \to +\infty$, to the point $z = 0$, so that all S_m are contained in a sufficiently large circular disk, and

[17] Jessen and Wintner [10], Theorem 5.

[18] *Ibid.*, Theorem 1 ($p = 1$).

so the convergence of (25) implies [19] the convergence of (31). This clearly completes the proof of the facts stated above.

In the cases to be considered in the following Sections, the infinite convolution (25) is convergent and is known [20] to be absolutely continuous with a density $\delta(x, y)$ which has continuous partial derivatives of arbitrarily high order, while the finite convolution (22) is absolutely continuous with a density $\delta_n(x, y)$ which has [21] continuous partial derivatives of order k for every $n \geqq 2k + 5$. This implies [22] that

$$\delta(x, y) = \lim_{n=\infty} \delta_n(x, y) \tag{32}$$

for every (x, y) in the z-plane. Finally, all the densities $\delta_n(x, y)$ occurring in (32) may be obtained from the density $\delta_2(x, y)$ discussed in Section 3 by means of the recursion formula

$$\delta_{n+1}(x, y) = \frac{1}{2\pi} \int_0^{2\pi} \delta_n(x - \xi_{n+1}(\theta), y - \eta_{n+1}(\theta))\, d\theta. \tag{33}$$

This is seen from (21) and from the definition of the corresponding distribution function $\phi_m(E)$, since $\psi_{n+1} = \psi_n * \phi_{n+1}$ by (22).

The spectrum (23) of (22) is [23] either a closed bounded region bordered by a convex curve C_n or a closed bounded ring-shaped region bordered by two disjoint convex curves C_n, D_n, where D_n will denote the inner boundary of (23). Similarly,[24] if (25) is convergent, then its spectrum is the infinite vectorial sum

$$V = S_1 \, (+) \, S_2 \, (+) \cdots \tag{34}$$

which, when a bounded set, is either a closed convex region bordered by a convex curve C or a ring-shaped region bordered by two disjoint convex curves C, D, where D will denote the inner boundary of (34). The rôle of the assumption (24) made in the above geometrical criterion for the convergence of (25) is that if (25) is convergent, then its spectrum (34) is a bounded

[19] *Ibid.*, Theorem 5.

[20] Wintner [16], pp. 328-329; Jessen and Wintner [10], Sections 7 and 13 and p. 53; Haviland and Wintner [9].

[21] *Ibid.*

[22] Wintner [18].

[23] Bohr [2]; Jessen and Wintner [10], Theorem 17.

[24] Cf. Bohr [2]; Jessen and Wintner [10], Theorems 3 and 17.

set[25] if and only if the diameters d_m of the spectra S_m of the distribution functions ϕ_m satisfy the condition (24). This condition for the boundedness of (34) will be satisfied in the cases to be considered, so that the density (32) of (25) cannot be regular analytic for every (x, y); cf. the Introduction.

In order to obtain, in the case belonging to (1), subregions of the spectrum (34) of (25) such that the density (32) of (25) is a regular analytic function of the two real variables x, y within each of these subregions, the results of Section 3 will be combined in Section 7 with the method previously applied[26] to the case of circular equidistribution, i. e., to the case which belongs to almost periodic functions with linearly independent frequencies. Since the proof is, in the case which belongs to (1), quite involved, first the known case of circular equidistributions will be treated by means of the direct method instead of the reduction to the one-dimensional case previously applied.[26]

5. *Almost periodic functions with linearly independent frequencies.* The asymptotic distribution function of the almost periodic function

$$f(t) = \sum_{m=1}^{\infty} a_m \exp i\lambda_m(t + \alpha_m), \text{ where } a_m > 0, \tag{35}$$

is[27] in case of linearly independent frequencies λ_m the infinite convolution (25), where $\phi_m(E)$ is the distribution function belonging to (21), if

$$S_m : x + iy = z = \xi_m(\theta) + i\eta_m(\theta) \equiv a_m \exp i\theta. \tag{36}$$

Since (35) is almost periodic with linearly independent frequencies, the series $\sum_{m=1}^{\infty} a_m$, where $a_m > 0$, is convergent by a well-known theorem of Bohr. Put

$$r_h = \sum_{m=h+1}^{\infty} a_m, \qquad (h = 0, 1, 2, \cdots), \tag{37}$$

and let, without loss of generality,

$$a_m \geqq a_{m+1} > 0. \tag{38}$$

Then the spectrum (34) of (25) is seen to be the circle $|z| = r_0$ or the circular ring $2a_1 - r_0 \leqq |z| \leqq r_0$ according as $2a_1 - r_0 = 0$ or $2a_1 - r_0 > 0$. The recursion formula (33) becomes

[25] This is clear by combining the remark of footnote [18] with Theorem 3 of Jessen and Wintner [10]. Notice that Theorem 18 given there assumes that the convex curves S_n surround the point $z = 0$.

[26] Kershner and Wintner [14].

[27] Jessen and Wintner [10], Section 13.

$$(39) \qquad \delta_{n+1}(x, y) = \frac{1}{2\pi} \int_0^{2\pi} \delta_n(x - a_{n+1} \cos \theta, y - a_{n+1} \sin \theta) \, d\theta$$

in view of (36).

It is easily verified that the open set R defined by (12) is, in the cas where S_1, S_2 are defined by (36), the punctured circle

$$(40a) \qquad 0 < |z| < a_1 + a_2, \qquad (z = x + iy),$$

or the circular ring

$$(40b) \qquad a_1 - a_2 < |z| < a_1 + a_2, \qquad (z = x + iy),$$

according as $a_1 = a_2$ or $a_1 > a_2$, the sets $T(\Omega^+)$, $T(\Omega^-)$ being the set $|z| = a_1 + a_2$, $|z| = a_1 - a_2$ in both cases. The representation (17) of th density $\delta_2(x, y)$ of (6) within R becomes

$$(40) \quad \delta_2(x, y) = \{\pi^2[(a_1 + a_2)^2 - (x^2 + y^2)][(x^2 + y^2) - (a_1 - a_2)^2]\}^{-1}$$

Now suppose that

$$(41_2) \qquad a_2 > r_2.$$

For a given positive $\epsilon < a_2 - r_2$, choose a positive $\epsilon = \bar{\epsilon}(\epsilon)$ so small that o: replacing (x, y) in (40) by $(x + iu, y + iv)$, where u, v are real, one obtain a function

$$(42_2) \qquad \delta_2(x + iu, y + iv)$$

which is regular analytic and bounded in the open set

$$(43_2) \qquad G_2 : \begin{cases} a_1 - a_2 + \epsilon < (x^2 + y^2)^{\frac{1}{2}} < a_1 + a_2 - \epsilon, \\ 0 \leqq (u^2 + v^2)^{\frac{1}{2}} < \bar{\epsilon} \end{cases}$$

of the two complex variables $x + iu$, $y + iv$. Since $\epsilon < a_2 - r_2 > 0$, it i clear from the definition (37) of r_n that

$$(43_n) \quad G_n : \begin{cases} a_1 - a_2 + \epsilon + \sum_{m=3}^{n} a_m < (x^2 + y^2)^{\frac{1}{2}} < a_1 + a_2 - \epsilon - \sum_{m=3}^{\infty} a_m, \\ 0 \leqq (u^2 + v^2)^{\frac{1}{2}} < \bar{\epsilon} \end{cases}$$

defines, for every $n \geqq 3$, an non-empty open set in the space of the two comple variables $x + iu$, $y + iv$. Let ϵ, $\bar{\epsilon}$ be fixed and let M denote a bound of th function (42_2) in the domain (43_2). Now it is clear from definitions (43_2) (43_n) that if G_{n+1}, where $n \geqq 2$, contains the point

$$(x + iu, y + iv), \text{ then } (x + iu - a_{n+1} \cos \theta, y + iv - a_{n+1} \sin \theta)$$

is a point of G_n for every θ. Hence, if there has been defined on G_n a regular analytic function

$$(42_n) \qquad \delta_n(x+iu, y+iv)$$

of $x+iy$, $u+iv$ such that $|\delta_n| \leqq M$ in G_n, then the formula

$$(44) \quad \delta_{n+1}(x+iu, y+iv) = \frac{1}{2\pi}\int_0^{2\pi} \delta_n(x+iu-a_{n+1}\cos\theta, y+iv-a_{n+1}\sin\theta)\,d\theta$$

defines on G_{n+1} a regular analytic function which has, in G_{n+1}, an absolute value

$$|\delta_{n+1}| \leqq \frac{1}{2\pi}\int_0^{2\pi} M\,d\theta = M.$$

Now if $u=0$, $v=0$, then (44) goes over into the representation (39) of the density $\delta_{n+1}(x,y)$ of ψ_{n+1}. Furthermore, the four-dimensional open set

$$(43_\infty) \quad G_\infty : \begin{cases} a_1 - a_2 + \epsilon + r_2 < (x^2+y^2)^{\frac{1}{2}} < a_1 + a_2 - \epsilon - r_2, \\ 0 \leqq (u^2+v^2)^{\frac{1}{2}} < \epsilon \end{cases}$$

where $0 < \epsilon < a_2 - r_2$ by assumption, is, in view of (37), clearly contained in every G_n, where $n \geqq 2$. Finally, (32) holds for every real (x,y), and so, in particular, at every real point $(u=0, v=0)$ of the domain G_∞ defined by (43_∞). It follows, therefore, from the extension of Vitali's theorem to uniformly bounded sequences of regular functions of two variables that if $n \to +\infty$ and $(x+iu, y+iv)$ is in G_∞, then the functions (42_n) tend to a function $\delta(x+iu, y+iv)$ which is regular analytic in G_∞ and reduces, if $u=0$, $v=0$, to the density $\delta(x,y)$ of the infinite convolution (25). Since $\epsilon > 0$ in (43_∞) is arbitrarily small, it follows that the density (32) of (25) is a regular analytic function of the two real variables x, y in the ring

$$(45_2) \qquad a_1 - a_2 + r_2 < (x^2+y^2)^{\frac{1}{2}} < a_1 + a_2 - r_2,$$

whenever (41_2) is satisfied.

It is clear that if the assumption (41_2) is replaced by

$$(41_k) \qquad a_k > r_k,$$

where $k \geqq 2$ has a fixed value, an obvious modification of the above proof yields that the density (32) of (25) is a regular analytic function of the two real variables x, y in the ring

$$(45_k) \qquad \sum_{m=1}^{k-1} a_m - a_k + r_k < (x^2 + y^2)^{\frac{1}{2}} < \sum_{m=1}^{k-1} a_m + a_k - r_k.$$

Furthermore, it is easily seen by the same argument that if (41_k) holds not only for a fixed $k \geqq 2$ but for $k = 1, \cdots, l$, where $l \geqq 2$ is fixed, then the density (32) of (25) is a regular analytic function of the two variables x, y not only within each of the $l - 1$ rings defined by (45_k) for $k = 2, \cdots, l$, but also within each of the $2^{l-1} - 1$ mutually disjoint rings

$$(46) \quad \sum_{m=1}^{k-1} e_m a_m - a_k + \sum_{m=k+1}^{\infty} a_m < (x^2 + y^2)^{\frac{1}{2}} < \sum_{m=1}^{k-1} e_m a_m + a_k - \sum_{m=k+1}^{\infty} a_m,$$

where $\sum_{m=k+1}^{\infty} a_m = r_k$ and

(47) $\quad e_1 = 1$, while $e_m = \pm 1$ for $m = 2, 3, \cdots, k - 1$; and $k = 2, 3, \cdots, l$,

it being understood that, since $e_1 = 1$, there are 2^{k-2} symbols $(e_1, \cdots, e_{k-1})$ for a fixed $k \geqq 2$.

Let it be mentioned for later reference that the a_m occurring in (46) are the supporting functions of the convex curves S_m, the latter being the circles $| z | = a_m$.

6. *The logarithmical derivative of the Riemann zeta function.* If $\sigma > 1$ is fixed, $s = \sigma + it$ and p_m is the m-th prime number, then the asymptotic distribution function of the almost periodic function

$$(48) \qquad f_\sigma(t) \equiv -\frac{\zeta'(s)}{\zeta(s)} + \text{const.} = \sum_{m=1}^{\infty} \frac{\log p_m}{p_m{}^s - 1} + \text{const.},$$

where const. $= \zeta'(2\sigma)/\zeta(2\sigma)$, may be represented [27] as the infinite convolution

$$(49) \qquad \psi^\sigma(E) = \psi^\sigma = \lim_{n=\infty} \psi_n{}^\sigma = \phi_1{}^\sigma * \phi_2{}^\sigma * \cdots,$$

where

$$(50) \qquad \psi_n{}^\sigma(E) = \psi_n{}^\sigma = \phi_1{}^\sigma * \phi_2{}^\sigma * \cdots * \phi_n{}^\sigma$$

and $\phi_m{}^\sigma = \phi_m{}^\sigma(E)$ denotes the distribution function which belongs to the curve (21) defined by

$$(51) \qquad S_m : x + iy = \xi_m(\theta) + i\eta_m(\theta) \equiv \frac{\log p_m}{p_m{}^\sigma e^{i\theta} - 1} - \frac{\log p_m}{p_m{}^{2\sigma} - 1}.$$

Thus S_m is the circle $| z | = a_m{}^\sigma$, where

$$(52) \qquad a_m{}^\sigma = \frac{p_m{}^\sigma \log p_m}{p_m{}^{2\sigma} - 1}, \qquad (a_m{}^\sigma > a_{m+1}{}^\sigma),$$

so that, since $\sigma > 1$, the series $\sum_{m=1}^{\infty} a_m^{\sigma}$ is convergent. Put

$$(53) \qquad r_h^{\sigma} = \sum_{m=h+1}^{\infty} a_m^{\sigma}, \qquad (h = 0, 1, \cdots),$$

and let σ_k be the obviously unique number such that

$$(54) \qquad a_k^{\sigma} = r_k^{\sigma} \text{ if } \sigma = \sigma_k > 1, \qquad (k = 1, 2, \cdots),$$

so that

$$(55) \qquad a_k^{\sigma} > r_k^{\sigma} \text{ if and only if } \sigma > \sigma_k.$$

According to the rule (34), the spectrum of the infinite convolution (49) is the infinite vectorial sum

$$(56) \qquad V^{\sigma} = S_1^{\sigma} \, (+) \, S_2^{\sigma} \, (+) \cdots .$$

It follows[28] that the spectrum of (49) is the circle $|z| \leqq r_0^{\sigma}$ or the circular ring $2a_1^{\sigma} - r_0^{\sigma} \leqq |z| \leqq r_0$ according as $\sigma \leqq \sigma_1$ or $\sigma > \sigma_1$, where $\sigma_1 = 2.576 \cdots$ by (54).

Now let, for a fixed integer $l \geqq 2$,

$$(57) \qquad \sigma > \mathrm{Max}\,(\sigma_1, \cdots, \sigma_l).$$

Then[29] the density $\delta^{\sigma} = \delta^{\sigma}(x, y)$ of the distribution function (49) is a regular analytic function of the two real variables x, y within each of the $2^{l-1} - 1$ mutually disjoint rings which result from (46), (47) by writing a_m^{σ} instead of a_m in (46).

In fact, the considerations of Section 5 may be repeated without any change if one proves that the density $\delta_2^{\sigma}(x, y)$ of $\phi_1^{\sigma} * \phi_2^{\sigma}$ is a regular analytic function of the two real variables x, y within the ring

$$a_1^{\sigma} - a_2^{\sigma} < (x^2 + y^2)^{\frac{1}{2}} < a_1^{\sigma} + a_2^{\sigma}.$$

Now this ring is, corresponding to (40b) and (55), identical with the open set R on which (17) is valid; and it is clear from (51) that (17) represents a regular analytic function on R. In fact, the functions (14) are regular analytic on R, since the Jacobian (8) vanishes only on Ω and the open set (12) does not contain points of the T-image of Ω.

---

[28] Bohr [1], Burrau [7].

[29] It will not be discusssed which of the l absolute constants $\sigma_1, \cdots, \sigma_l$ is Max $(\sigma_1, \cdots, \sigma_l)$ for a fixed l and what is the asymptotic behavior of σ_l and Max $(\sigma_1, \cdots, \sigma_l)$ for large l.

7. *The logarithm of the Riemann zeta function.* If $\sigma > 1$ is fixed, $s = \sigma + it$ and p_m is the m-th prime number, then the asymptotic distribution function of the almost periodic function

$$(58) \qquad f_\sigma(t) \equiv -\log \zeta(s) + \text{const.} = \sum_{m=1}^{\infty} \log(1 - p_m{}^{-s}) + \text{const.},$$

where const. $= \frac{1}{2} \log \zeta(2\sigma)$, may be represented [27] as the infinite convolution $\psi^\sigma(E)$ defined by (49) and (50), where $\phi_m{}^\sigma = \phi_m{}^\sigma(E)$ denotes the distribution function which belongs to the curve (21) defined by

$$(59) \quad S_m{}^\sigma : x + iy = \xi_m(\theta) + i\eta_m(\theta)$$
$$= \log(1 - p_m{}^{-\sigma} e^{i\theta}) - \tfrac{1}{2}\log(1 - p_m{}^{-2\sigma}) \equiv g(\theta; p_m{}^{-\sigma}) + ih(\theta; p_m{}^{-\sigma});$$
$$g(\theta;\rho) = \tfrac{1}{2}\log\frac{1 - 2\rho\cos\theta + \rho^2}{1 - \rho^2}, \qquad h(\theta;\rho) = \arctan\frac{\rho\sin\theta}{1 - \rho\cos\theta},$$
$$|h(\theta;\rho)| < \tfrac{1}{2}\pi.$$

Thus S_m is a regular analytic convex curve which has both axes $x = 0$, $y = 0$ as lines of symmetry.[30]

Let $a_m{}^\sigma(\omega)$ denote the supporting function (Stützfunktion) of $S_m{}^\sigma$, i. e., the distance of the origin $z = 0$ from that oriented normal of $S_m{}^\sigma$ which has the inclination ω. Then [30]

$$(60) \qquad \begin{aligned} a_m{}^\sigma(\tfrac{1}{2}\pi) &= \arcsin p_m{}^{-\sigma} (< \tfrac{1}{2}\pi), \\ a_m{}^\sigma(0) &= \tfrac{1}{2}\log\left[(1 + p_m{}^{-2\sigma})^2/(1 - p_m{}^{-2\sigma})\right], \end{aligned}$$

while

$$(61) \qquad 0 < a_m{}^\sigma(\omega) < a_m{}^\sigma(0) + a_m{}^\sigma(\tfrac{1}{2}\pi) \text{ for } 0 \leqq \omega < 2\pi,$$

since the two-fold symmetry of the convex curve $S_m{}^\sigma$ implies that

$$a_m{}^\sigma(\omega) = a_m{}^\sigma(\omega + \pi), \qquad a_m{}^\sigma(\omega) = a_m{}^\sigma(\pi - \omega).$$

Since $\sigma > 1$, it follows that the series

$$(62) \qquad r_h{}^\sigma(\omega) = \sum_{m=h+1}^{\infty} a_m{}^\sigma(\omega), \qquad (h = 0, 1, \cdots),$$

are uniformly convergent for $0 \leqq \omega < 2\pi$. It is known [31] that if σ and k are fixed, then the inequality $a_k{}^\sigma(\omega) > r_k{}^\sigma(\omega)$ holds for every ω if it holds for $\omega = \frac{1}{2}\pi$. Hence, if $\sigma^{(k)}$ denotes, for a fixed positive integer k, the obviously unique number such that

[30] Cf. Bohr and Courant [4] or Bohr and Jessen [6].
[31] Bohr and Jessen [6], p. 40.

(63) $\qquad \arcsin p_k^{-\sigma} = \sum_{m=k+1}^{\infty} \arcsin p_m^{-\sigma}$ if $\sigma = \sigma^{(k)} > 1, \qquad (k = 1, 2, \cdots),$

then it is seen from (60) that

(64) $\qquad \sigma > \sigma^{(k)}$ if and only if $a_k^{\sigma}(\omega) > r_k^{\sigma}(\omega)$ for all ω,

where k is fixed.

The spectrum (56) of the infinite convolution (49) belonging to the curves (59) may be described as follows: The spectrum V^{σ} is [32] a closed bounded region bordered by a convex curve C^{σ} or a closed bounded ring-shaped region bordered by two disjoint convex curves C^{σ}, D^{σ} according as $1 < \sigma \leqq \sigma^{(1)}$ or $\sigma > \sigma^{(1)}$, where $\sigma^{(1)} = 1.764 \cdots$ is defined by (63). If $\sigma > \sigma^{(1)}$, i. e., if D^{σ} exists, let C^{σ} denote that of the two convex curves C^{σ}, D^{σ} which forms the outer boundary of V^{σ}. Let $\bar{\sigma}$ denote the obviously unique root of the equation

(65) $$p_1^{-\sigma} = \sum_{m=2}^{\infty} p_m^{-\sigma}$$

in the range $\sigma > 1$, so that [33] $\bar{\sigma} = 1.778 \cdots$, and so $\bar{\sigma} < \sigma^{(1)}$. Then, if $\sigma > \sigma^{(1)}$, i. e., if D^{σ} exists, the symmetric convex curve D^{σ} has [34] no corners or has corners on the x-axis and nowhere else according as $\sigma \geqq \bar{\sigma}$ or $\sigma < \bar{\sigma}$. Furthermore, if $\sigma \geqq \bar{\sigma}$, then the supporting function of D^{σ} is [35]

(66) $$D^{\sigma} : a_1^{\sigma}(\omega) - r_1^{\sigma}(\omega), \qquad (\sigma \geqq \sigma).$$

The supporting function of C^{σ} is [36] $r_0^{\sigma}(\omega)$ for every $\sigma > 1$.

Let

(67) $\qquad e_1 = 1$ and $e_m = \pm 1$, where $m = 2, 3, \cdots,$

so that there is a non-enumerable set of symbols $(e_1, e_2, \cdots)$. The relations (60) and (61) imply that the series

(68) $$r^{\sigma}(\omega; \{e_m\}) = \sum_{m=1}^{\infty} e_m a_m^{\sigma}(\omega), \qquad (\sigma > 1),$$

is, for every symbol

(69) $$(e_1, e_2, \cdots, e_m, \cdots) = (1, e_2, \cdots, e_m, \cdots),$$

[32] Bohr and Jessen [6].

[33] Bohr and Jessen [6], p. 43.

[34] Cf. Bohr and Jessen [6], p. 43 and Kershner [12].

[35] Kershner [11], where the result is proved for an arbitrary finite number of convex curves, but the proof holds for the infinite vectorial sum also.

[36] Haviland [8].

uniformly convergent for $0 \leqq \omega < 2\pi$. It is clear from (67), (68) and from the definition (62) of $r_1^\sigma(\omega)$ that $r^\sigma(\omega; \{e_m\})$ is the sum of $a_1^\sigma(\omega) - r_1^\sigma(\omega)$ and of a finite or infinite series the terms of which are some of the functions $a_m^\sigma(\omega) > 0$. This holds for every $\sigma > 1$. Now if $\sigma \geqq \bar{\sigma}\ (> 1)$, then $a_1^\sigma(\omega) - r_1^\sigma(\omega)$ is, by (66), the supporting function of the convex curve D^σ, while $a_m^\sigma(\omega)$ is the supporting function of the convex curve S_m^σ for any $\sigma > 1$. Hence the function (68) of ω is, by a classsical theorem,[37] the supporting function of a convex curve for every $\sigma \geqq \sigma$. This convex curve, which depends on the symbol (69), will be denoted by

$$(70) \qquad C^\sigma(\{e_m\}), \qquad (\sigma \geqq \bar{\sigma}).$$

Now let,[38] for a fixed integer $l \geqq 2$,

$$(71) \qquad \sigma > \text{Max}\,(\sigma^{(1)}, \cdots, \sigma^{(l)})$$

and choose, for a fixed integer k which satisfies the inequalities $2 \leqq k \leqq l$, a pair of symbols (69) which will be denoted by

$$(72) \qquad \{e_m\}_k^{\mathrm{I}}, \quad \{e_m\}_k^{\mathrm{II}}, \qquad (2 \leqq k \leqq l),$$

and have the property that the numbers $e_1 = 1, e_2, \cdots, e_{k-1}$ are respectively the same in both symbols, while $e_k = 1, e_{k+1} = -1, e_{k+2} = -1, e_{k+3} = -1, \cdots$ in $\{e_m\}_k^{\mathrm{I}}$ and $e_k = -1,\ e_{k+1} = 1,\ e_{k+2} = 1,\ e_{k+3} = 1, \cdots$ in $\{e_m\}_k^{\mathrm{II}}$. Thus there are 2^{k-2} pairs of symbols (72) for a fixed k, and so $2^{l-1} - 1$ pairs of symbols (72) for $2 \leqq k \leqq l$. Now if (72) is any of these $2^{l-1} - 1$ pairs of symbols and if (71) is satisfied, then the convex curve (70) which belongs to $\{e_m\} = \{e_m\}_k^{\mathrm{II}}$ lies entirely within the convex curve (70) which belongs to $\{e_m\} = \{e_m\}_k^{\mathrm{I}}$. In fact, it is clear from the definition of the convex curve (70) that the difference of the supporting functions of the convex curves

$$(73\text{a}) \qquad C^\sigma(\{e_m\}_k^{\mathrm{I}}) \text{ and } C^\sigma(\{e_m\}_k^{\mathrm{II}})$$

is $2a_k^\sigma(\omega) - 2r_k^\sigma(\omega)$, an expresssion which is positive for every ω in view of (64) and (71). This clearly [37] proves the statement. The *open* ring-shaped domain bordered by the two convex curves (73a) will be denoted by

$$(73\text{b}) \qquad H^\sigma(\{e_m\}_k), \qquad (2 \leqq k \leqq l).$$

---

[37] Cf., e. g., Haviland [8].

[38] It will not be discussed which of the l absolute constants $\bar{\sigma}, \sigma^{(2)}, \cdots, \sigma^{(l)}$ is Max $(\bar{\sigma}, \sigma^{(1)}, \cdots, \sigma^{(l)})$ for a fixed l and what is the asymptotic behaviour of $\sigma^{(l)}$ and Max $(\bar{\sigma}, \sigma^{(1)}, \cdots, \sigma^{(l)})$ for large l.

It is seen by an obvious repetition of the above consideration that none of the infinitely many convex curves (70) has any point within the open set (73b). This implies that the $2^{l-1}-1$ ring-shaped domains (73a) which belong to all pairs of symbols (72) are mutually disjoint.

Now it is easily seen that if $\delta^{\sigma}(x,y)$ denotes the density of the infinite convolution (49) in which $\phi_m{}^{\sigma}$ belongs to the curve (59), and if σ is chosen such that (71) is satisfied, then $\delta^{\sigma}(x,y)$ is a regular analytic function of the two real variables x, y within each of the $2^{l-1}-1$ mutually disjoint ring-shaped open sets (73b).

First, let $\delta_2{}^{\sigma}(x,y)$ denote the density of the convolution $\psi_2{}^{\sigma}=\phi_1{}^{\sigma}*\phi_2{}^{\sigma}$ (cf. Section 3), where $\phi_1{}^{\sigma}$, $\phi_2{}^{\sigma}$ are the distribution functions belonging to the curves $S_1{}^{\sigma}$, $S_2{}^{\sigma}$ defined by (59). Since (71) implies that $\sigma\geqq\bar{\sigma}$, the difference $a_1{}^{\sigma}(\omega)-r_1{}^{\sigma}(\omega)$ is the supporting function of a convex curve; cf. (66). Since the function $r_2{}^{\sigma}(\omega)$ defined by (62) also is[37] the supporting function of a convex curve, it follows[37] that the same holds for the sum of $a_1{}^{\sigma}(\omega)-r_1{}^{\sigma}(\omega)$ and $r_2{}^{\sigma}(\omega)$. This means in view of (62) that

$$a_1{}^{\sigma}(\omega)-a_2{}^{\sigma}(\omega) \tag{74}$$

is the supporting function of a convex curve. It follows[39] that the vectorial sum $S_1{}^{\sigma}\ (+)\ S_2{}^{\sigma}$ is ring-shaped, with an inner boundary curve $D_2{}^{\sigma}$ which is free of corners and has the difference (74) as its supporting function. Hence it is seen from the last remark in Section 2 that the open set $R=R^{\sigma}$ on which the explicit representation (17) of $\delta_2=\delta_2{}^{\sigma}$ is valid is identical with the open ring-shaped domain bordered by the two convex curves $C_2=C_2{}^{\sigma}$, $D_2=D_2{}^{\sigma}$, i.e., by the convex curves which have the supporting functions

$$a_1{}^{\sigma}(\omega)+a_2{}^{\sigma}(\omega);\qquad a_1{}^{\sigma}(\omega)-a_2{}^{\sigma}(\omega). \tag{75}$$

This open ring-shaped domain will be denoted by $H_2{}^{\sigma}$, so that

$$H_2{}^{\sigma}=[S_1{}^{\sigma}\ (+)\ S_2{}^{\sigma}]-[C_2{}^{\sigma}+D_2{}^{\sigma}]. \tag{76}$$

Now the functions $\xi_m(\theta)$, $\eta_m(\theta)$ occurring in the definition (59) of $S_m{}^{\sigma}$ are regular analytic. Furthermore, the Jacobian (8) of the transformation (7), when expressed by means of (14) in terms of x, y, does not vanish on the set $H_2{}^{\sigma}$ (cf. the corresponding argument at the end of Section 6). Since the representation (17) of $\delta_2=\delta_2{}^{\sigma}$ is valid on $H_2{}^{\sigma}$, it follows that $\delta_2{}^{\sigma}$ is a

[39] Kershner [11], Theorem III.

regular analytic function of the two real variables x, y on the open ring-shaped domain H_2^σ.

Let $\delta_n^\sigma(x, y)$ denote the density of the distribution function (50), so that, by (33) and (59),

$$(77)\quad \delta_n^\sigma(x, y) = \frac{1}{2\pi}\int_0^{2\pi} \delta^\sigma_{n-1}(x - g(\theta; p_n^{-\sigma}), y - h(\theta; p_n^{-\sigma}))\, d\theta, \text{ where } n \geqq 3$$

It is obvious from the same consideration which has been applied before to H_2^σ that both functions

$$(78)\qquad a_1^\sigma(\omega) + a_2^\sigma(\omega) - \sum_{m=3}^{\infty} a_m^\sigma(\omega); \qquad a_1^\sigma(\omega) - a_2^\sigma(\omega) + \sum_{m=3}^{\infty} a_m^\sigma(\omega)$$

are supporting functions of certain convex curves which form the outer and inner boundary of an open ring-shaped domain. On denoting this domain by H_n^σ, it is clear that H_{n+1}^σ is a subset of H_n^σ and that, for a fixed σ, the common part of all H_n^σ is the unique ring-shaped region (73b) which belongs to $k = 2$. Since the two functions (78) represent the supporting functions of the outer and inner boundaries of H_n^σ, it follows that if H_n^σ contains the point (x, y), then

$$(x - g(\theta; p_n^{-\sigma}), y - h(\theta; p_n^{-\sigma}))$$

is a point of H_{n-1}^σ for every θ. Hence the argument which led from (39) to (45_2) can be applied without any modification and shows that the density $\delta^\sigma(x, y)$ of (49) is, again in view of (32) and by virtue of Vitali's theorem, a regular analytic function of the two real variables x, y on the ring-shaped region (73b) which belongs to $k = 2$. Finally, the regular analyticity of $\delta^\sigma(x, y)$ in the remaining $2^{l-1} - 1 - 1$ open ring-shaped regions (73_2) follows in the same way; cf. the end of Section 5.

Appendix.[40] *On real almost periodic functions with linearly independent frequencies.* It is known [41] that if

$$(1)\qquad f(t) \sim \sum_{n=0}^{\infty} a_n \cos(\lambda_n t + \delta_n), \qquad (a_n > 0),$$

is almost periodic, then it has an asymptotic distribution function $\sigma(x)$, and that the spectrum of $\sigma(x)$ consists of the closure of the set of values attained

[40] The numbers of the formulae will refer to the formulae of this Appendix.

[41] Wintner [19], pp. 254-255 and p. 269; cf. also Wintner [20].

by $x = f(t)$ for $-\infty < t < +\infty$. In what follows, it will be assumed that the frequencies λ_k of (1) are linearly independent. Then the series

$$r = \sum_{n=0}^{\infty} a_n, \qquad (a_n > 0), \tag{2}$$

is convergent by a well-known theorem of Bohr, and the closure just mentioned is the interval $-r \leqq x \leqq r$. Furthermore, on placing

$$g(x) = \tfrac{1}{\pi}(1 - x^2)^{-\frac{1}{2}}, \text{ if } |x| < 1; \quad g(x) = 0, \text{ if } |x| \geqq 1, \tag{3}$$

and defining a sequence $\sigma_0, \sigma_1, \cdots$ of distribution functions by means of the recursion formula

$$\sigma_n(x) = \int_{-\infty}^{+\infty} \sigma_{n-1}(x - a_n y) g(y)\, dy, \tag{4}$$

where

$$\sigma_0(x) = \int_{-\infty}^{x} g(y/a_0)\, dy/a_0, \tag{5}$$

it is known [42] that the asymptotic distribution function $\sigma(x)$ of (1) is represented for every x by

$$\sigma(x) = \lim_{n=\infty} \sigma_n(x), \tag{6}$$

and that [43] every derivative

$$\sigma^{(m)}(x) \text{ exists for } m = 1, 2, \cdots; \; -\infty < x < +\infty. \tag{7}$$

Also, $\sigma_n(x)$ has [43] a continuous m-th derivative for every x, whenever

$$n \geqq 2m + 2. \tag{8}$$

Finally, as will be shown at the end of this paper,

$$\lim_{n=\infty} \sigma_n^{(m)}(x) = \sigma^{(m)}(x); \; -\infty < x < +\infty, \qquad (m = 1, 2, \cdots). \tag{9}$$

Since the spectrum of $\sigma(x)$ is the interval $|x| \leqq r$, it is clear that $\sigma(x)$ cannot be regular analytic at $x = \pm r$. On the other hand, it is known [44]

[42] Wintner [16], pp. 312-316.
[43] Wintner [16], p. 315.
[44] Kershner and Wintner [14].

that if the coefficients a_k of (1) satisfy certain inequalities, one can delimit within the spectrum $|x| \leqq r$ subregions of regular analyticity of $\sigma(x)$.

It was not proved so far that the end-points of the subregions determined in the paper just mentioned are singularities of $\sigma(x)$. There was not even known an example with any established singularity for $\sigma(x)$ in the interior of the spectrum. The object of the following considerations is to fill into these gaps. The method to be applied also gives some qualitative information as to the shape of the density curve $\sigma'(x)$; cf. (ii) and (iv) below. Needless to say, all singularities of $\sigma(x)$ are essential singularities in virtue of (7).

The results to be obtained may be summarized as follows:

(i) If

$$r < 2a_0, \tag{10}$$

then $\sigma(x)$ is regular analytic in the subinterval

$$-(2a_0 - r) < x < (2a_0 - r) \tag{11}$$

of the spectrum $|x| \leqq r$ and has at the end-points $x = \pm(2a_0 - r)$ of (11) essential singularities.

(ii) If (10) is satisfied, there exists for every $m \geqq 1$ an $\epsilon_m > 0$ such that the derivative $\sigma^{(m)}(x)$ is positive for $0 < x < 2a_0 - r + \epsilon_m$. On choosing $m = 2$ and $m = 3$, it follows that there exists an $\epsilon > 0$ such that the density $\sigma'(x)$ is monotone increasing and convex in the interval $0 \leqq x \leqq 2a_0 - r + \epsilon$.

(iii) If

$$r = 2a_0 \tag{12}$$

then $\sigma(x)$ has an essential singularity at $x = 0$.

(iv) If (12) is satisfied, there exists for every $m \geqq 1$ an $\epsilon_m > 0$ such that the derivative $\sigma^{(m)}(x)$ is positive for $0 < x < \epsilon_m$. On choosing $m = 2$ and $m = 3$, it follows that there exists an $\epsilon > 0$ such that the density $\sigma'(x)$ is monotone increasing and convex in the interval $0 \leqq x \leqq \epsilon$.

(It may be mentioned that on combining (ii) and (iv) with the representation of $\sigma(x)$ which is obtained [45] by Fourier inversion, there results the following curiosity: If (2) is a convergent series with positive terms which satisfy the inequality $r \leqq 2a_0$, then the number I_m defined by the definite Bessel integral

[45] Wintner [16], p. 315.

$$I_m = \int_0^\infty x^{2m} \prod_{n=0}^{\infty} J_0(a_n x)\, dx$$

is positive or negative according as the positive integer m is even or odd.)

First, it is easily seen from (3) and (4) that if $n \geqq 1$ and $p \geqq 0$, then, for every x,

$$(13) \qquad \sigma_{p+n}(x) = \int_{-1}^{1} \cdots \int_{-1}^{1} \sigma_p(x - \sum_{i=1}^{n} a_{p+i} y_i)\, d\Omega,$$

where

$$(14) \qquad d\Omega = \prod_{i=1}^{n} [\tfrac{1}{\pi} (1 - y_i^2)^{-\frac{1}{2}} dy_i].$$

In particular,

$$(15) \qquad \sigma_n(x) = \int_{-1}^{1} \cdots \int_{-1}^{1} \sigma_0(x - \sum_{i=1}^{n} a_i y_i)\, d\Omega.$$

It is understood that the integrals are n-fold, so that, from (14),

$$(16) \qquad \int_{-1}^{1} \cdots \int_{-1}^{1} d\Omega = [\tfrac{1}{\pi} \int_{-1}^{1} (1 - y^2)^{-\frac{1}{2}} dy]^n = 1.$$

It is easily verified from (3), (5) and (15) that

$$\sigma_n(x) + \sigma_n(-x) = 1$$

for every n and x; hence

$$(17_n) \qquad \sigma_n^{(m)}(x) = (-1)^{m+1} \sigma_n^{(m)}(-x), \qquad (n \geqq 0,\ m \geqq 1),$$

for all those x at which σ_n has an m-th derivative.

The proof of (i) proceeds as follows. Suppose that (10) is satisfied. Then it is clear from (2) that

$$(18_n) \qquad -(a_0 - \sum_{i=1}^{n} a_i) < x < (a_0 - \sum_{i=1}^{n} a_i)$$

is an interval containing the interval (18_{n+1}) and the interval (11). It is also seen that if $p \geqq 0$ and $n \geqq 1$, and if x lies in the interval (18_{p+n}), then the number

$$(19) \qquad x - \sum_{i=1}^{n} a_{p+i} y_i$$

lies in the interval (18_n) for all those $(y_1, \cdots, y_n)$ for which $|y_i| < 1$, where $i = 1, \cdots, n$. On combining this fact with (15), it is seen from the definition (5), (3) of $\sigma_0(x)$ that

(20) $\sigma_n(x)$ is regular analytic in (18_n),

and that

$$(21_n) \qquad \sigma_n^{(m)}(x) = \int_{-1}^{1} \cdots \int_{-1}^{1} \sigma_0^{(m)}(x - \sum_{i=1}^{n} a_i y_i)\, d\Omega \text{ in } (18_n).$$

Now (21_n) means that

$$(22_n) \qquad \sigma_{p+n}^{(m)}(x) = \int_{-1}^{1} \cdots \int_{-1}^{1} \sigma_p^{(m)}(x - \sum_{i=1}^{n} a_{p+i} y_i)\, d\Omega \text{ in } (18_{p+n})$$

holds for $p = 0$. On combining (21_p) with (21_{p+n}), it is seen from the remark made in connection with (19) that (22_n) holds for every p.

It is known [46] that

(23) $\sigma(x)$ is regular analytic in (11).

In order to discuss the Taylor series of $\sigma(x)$ in the vicinity of $x = 0$, notice first that all coefficients in the binomial expansion

$$(24) \qquad (1 - x^2)^{-\frac{1}{2}} = \sum_{n=0}^{\infty} b_n x^{2n}, \qquad |x| < 1,$$

are positive, so that, from (3) and (5),

$$(25) \qquad \sigma_0^{(m)}(x) > 0, \text{ if } 0 < x < a_0.$$

It is also seen that

$$(26) \qquad \sigma_0^{(2q-1)}(0) > 0, \qquad (q \geqq 1),$$

while, from (17_n),

$$(27) \qquad \sigma_n^{(2q)}(0) = 0, \qquad (n \geqq 0,\ q \geqq 1).$$

Now (21_n), (25) and (17_0) clearly imply that

$$(28_q) \qquad \sigma_n^{(2q-1)}(x) > 0 \text{ in } (18_n).$$

It is seen from (27) and (28_{q+1}) that the minimum of the function $\sigma_n^{(2q-1)}(x)$ in the interval (18_n) is attained at $x = 0$, so that, for every $q \geqq 1$,

---

[46] Kershner and Wintner [14].

(29_n) $$\sigma_n^{(2q-1)}(x) \geqq \sigma_n^{(2q-1)}(0) \text{ in } (18_n).$$
On placing in (22_n)
$$x = 0 \text{ and } m = 2q - 1,$$

it is seen from (29_p) and from the remark made in connection with (19) that

$$\sigma_{p+n}^{(2q-1)}(0) > \sigma_p^{(2q-1)}(0) \int_{-1}^{1} \cdots \int_{-1}^{1} d\Omega.$$

Hence, from (16) and (28_q),

(30) $$\sigma_{p+n}^{(2q-1)}(0) > \sigma_p^{(2q-1)}(0) > 0.$$

Now from (9) and (30)

(31) $$\sigma^{(2q-1)}(0) > \sigma_p^{(2q-1)}(0) > 0,$$

while from (9) and (27)

(32) $$\sigma^{(2q)}(0) = \sigma_p^{(2q)}(0) = 0,$$
so that
(33) $$\sigma^{(m)}(0) \geqq 0$$
for every m.

Let R denote the radius of convergence of the power series

(34) $$\sigma(x) = \sum_{m=0}^{\infty} \sigma^{(m)}(0) x^m / m!$$

and R_n that of the power series

(35) $$\sigma_n(x) = \sum_{m=0}^{\infty} \sigma_n^{(m)}(0) x^m / m!\,.$$

On writing n instead of p in (31) and (32), it is seen that

(36) $$R \leqq R_n$$

for every n. On the other hand, on keeping n fixed and choosing m in (21_n) sufficiently large, it is seen from (3) and (5) that the regular analytic function representing $\sigma_n(x)$ in the interval (18_n) is singular at the end-points

$$x = \pm \left(a_0 - \sum_{i=1}^{n} a_i\right)$$

of (18_n), so that

(37) $$R_n \leqq a_0 - \sum_{i=1}^{n} a_i.$$

On letting $n \to \infty$ in (36) and (37), it follows from (2) that

$$(38) \qquad R \leqq 2a_0 - r; \text{ and } R > 0,$$

by (23). Now it is known that if the coefficient of a power series

$$(39) \qquad f(z) = \sum_{m=0}^{\infty} c_m z^m$$

are real and non-negative, and if this power series is convergent for $|z| < s$ but not for $|z| > s$, then $z = s > 0$ is a singular point of the function $f(z)$ also when the series (39) is convergent at $z = s$ (Vivanti-Pringsheim). It follows, therefore, from (33), (23) and (38) that $R = 2a_0 - r$ and that the function (34) is singular at $x = 2a_0 - r$, and so, by (32), at $x = -(2a_0 - r)$ also. This completes the proof of (i).

Since (34) is valid for $0 < x < 2a_0 - r$, so is the expansion

$$(40) \qquad \sigma^{(m)}(x) = \sum_{n=0}^{\infty} \sigma^{(n+m)}(0) x^n / n!.$$

On letting $x \to 2a_0 - r$ in (40), it is seen from (33) and (7) that, in view of Abel's continuity theorem, (40) is valid at $x = 2a_0 - r$ also. Consequently, $\sigma^{(m)}(x)$ is positive not only for $0 < x < 2a_0 - r$ but for $x = 2a_0 - r$ as well, and so, by (7), for $2a_0 - r < x < 2a_0 - r + \epsilon_m$ also, if $\epsilon_m > 0$ is sufficiently small. This proves (ii).

Now replace (10) by (12). It is clear that the proof of (31) and (32) is valid in the case (12) also, and that the radius of convergence R_n (> 0) of the power series (35) again satisfies the inequality (37). It is seen from (2), (12) and (37) that

$$(41) \qquad R_n \to 0 \text{ as } n \to \infty.$$

On combining (31) and (32) with (41), it follows that the power series (34) is divergent for every $x \neq 0$. This proves (iii).

Finally, the proof of (iv) is the same as that of (ii).

The statement (9) used above may be proved as follows. For a fixed $m \geqq 1$, let $\rho_0(x)$ denote the convolution of the $2m + 5$ distribution functions

$$\sigma_0(xa_0/a_j), \text{ where } 0 \leqq j \leqq 2m + 4,$$

so that $\rho_0(x)$ has, in view of the remark made in connection with (8), a continuous $(m+1)$-th derivative for every x. Furthermore, all derivatives of $\rho_0(x)$ vanish for sufficiently large $|x|$, since the spectrum of $\rho_0(x)$ is contained in the spectrum $|x| \leqq r$ of (6). Finally, on placing

$$\rho_n(x) = \sigma_0(xa_0/a_{n+2m+5}), \text{ where } n = 1, 2, \cdots,$$

the distribution function (6) is the infinite convolution of the distribution functions $\rho_0, \rho_1, \rho_2, \cdots$, since it is the infinite convolution of the distribution functions

$$\sigma_0(xa_0/a_k), \text{ where } k = 0, 1, 2, \cdots.$$

Hence (9) is implied by the following

LEMMA. If an infinite convolution of a sequence of distribution functions $\rho_0(x), \rho_1(x), \cdots$ converges to a distribution function $\sigma(x)$, and if $\rho_0(x)$ has for $-\infty < x < +\infty$ an absolutely integrable and bounded $(m+1)$-th derivative, then the convolution of the $n+1$ distribution functions $\rho_0(x), \cdots, \rho_n(x)$, has a continuous m-th derivative which tends, for $-\infty < x < +\infty$, to a continuous function representing the m-th derivative of $\sigma(x)$.

Now the proof of this Lemma requires but an obvious modification of the proof previously [47] given for the particular case $m = 1$.

THE JOHNS HOPKINS UNIVERSITY.

BIBLIOGRAPHY.

[1] H. Bohr, "Über die Funktion $\zeta'(s)/\zeta(s)$," *Journal für Mathematik*, vol. 141 (1912), pp. 217-234.

[2] ———, "Om Addition af uendelig mange konvekse Kurver," *Danske Videnskabernes Selskab*, Forhandlinger, 1913, pp. 325-366.

[3] ———, "Om Addition af konvekse Kurver med givne Sandsynlighedsfordelinger," *Matematisk Tidsskrift B*, 1923, pp. 10-15.

[4] ———, and R. Courant, "Neue Anwendungen der Theorie der diophantischen Approximationen auf die Riemannsche Zetafunktion," *Journal für Mathematik*, vol. 144 (1914), pp. 249-274.

[5] ———, and B. Jessen, "Om Sandsynlighedsfordelinger ved Addition af konvekse Kurver," *Danske Videnskabernes Selskab*, Skrifter (8), vol. 12, no. 3 (1929).

[6] ———, and B. Jessen, "On the distribution of the values of the Riemann zeta function," *American Journal of Mathematics*, vol. 58 (1936), pp. 35-44.

[7] C. Burrau, "Numerische Lösung der Gleichung . . . ," *Journal für Mathematik*, vol. 142 (1912), pp. 51-53.

[47] Wintner [18].

[8] E. K. Haviland, "On the addition of convex curves in Bohr's theory of Dirichlet series," *American Journal of Mathematics*, vol. 55 (1933), pp. 332-334.

[9] ———, and A. Wintner, "On the Fourier transforms of distributions on convex curves," *Duke Mathematical Journal*, vol. 2 (1936), pp. 712-721.

[10] B. Jessen and A. Wintner, "Distribution functions and the Riemann zeta function," *Transactions of the American Mathematical Society*, vol. 38 (1935), pp. 48-88.

[11] R. B. Kershner, "On the addition of convex curves," *American Journal of Mathematics*, vol. 58 (1936), pp. 737-746.

[12] ———, "On the values of the Riemann ζ-function on fixed lines $\sigma > 1$," *ibid.*, vol. 59 (1937), pp. 171-178.

[13] ———, and A. Wintner, "On symmetric Bernoulli convolutions," *ibid.*, vol. 57 (1935), pp. 541-548.

[14] ———, and A. Wintner, "On the asymptotic distribution of almost periodic functions with linearly independent frequencies," *ibid.*, vol. 58 (1936), pp. 91-94.

[15] ———, and A. Wintner, "On the boundary of the range of values of $\zeta(s)$," *ibid.*, vol. 58 (1936), pp. 421-425.

[16] A. Wintner, "Upon a statistical method in the theory of diophantine approximations," *ibid.*, vol. 55 (1933), pp. 309-331.

[17] ———, "On analytic convolutions of Bernoulli distributions," *ibid.*, vol. 56 (1934), pp. 659-663.

[18] ———, "On the differentiation of infinite convolutions," *ibid.*, vol. 57 (1935), pp. 363-366.

[19] ———, *Spektraltheorie der unendlichen Matrizen*, Leipzig, 1929.

[20] ———, "Über die Grundlagen des matrizenmechanischen Umdeutungsprinzips und über eine Spektraltheorie der Bohrschen fastperiodischen Funktionen," *Zeitschrift für Physik*, vol. 48 (1928), pp. 149-161.

ON THE ORDER OF THE COEFFICIENTS OF A UNIVALENT FUNCTION.[1]

By M. S. Robertson.

Introduction. If $f(z) = z + \sum_{2}^{\infty} a_n z^n$ is holomorphic and univalent for $|z| < 1$ then it is known[2] that $a_n = O(n)$. It has been conjectured[3] that $|a_n| \leqq n$. For $n = 2$ and 3 this conjecture is known to be true, equality occurring only for the function $z(1 + ze^{i\alpha})^{-2}$, α real. It is the purpose of the present paper to show that the univalent functions $f(z)$ for which

$$\limsup \frac{\log |a_n|}{\log n} = 1 \tag{1}$$

is true are also closely related to this same extremal function $z(1 + ze^{i\alpha})^{-2}$.

We show first of all that the only univalent function $f(z)$ for which

$$\left[f(x) - f\left(\frac{x - z}{1 - \bar{x}z}\right)\right] \cdot [(1 - |x|^2) f'(x)]^{-1} \equiv z(1 + ze^{i\alpha})^{-2} \tag{2}$$

for any x of modulus less than unity, and α a real number, is true is the very special function

$$f(z) \equiv \frac{z + (\bar{x} - xe^{2i\alpha})(\bar{x}e^{2i\alpha} - 1)^{-1} z^2}{\{1 - (\bar{x} + e^{i\alpha})(xe^{i\alpha} + 1)^{-1} z\}^2}. \tag{3}$$

Secondly, we show that a necessary condition that $f(z)$ satisfy (1) is that $f(z)$ be a rather special type of function satisfying (2) in an asymptotic sense. More precisely, in order that $f(z)$ should satisfy (1) it must necessarily possess the following very special property: there shall exist a sequence of points x_i inside the unit circle tending to a singularity of $f(z)$ on the circumference, and a real number α, so that

$$\lim_{|x_i| \to 1} \left[f(x_i) - f\left(\frac{x_i - z}{1 - \bar{x}_i z}\right)\right] \cdot [(1 - |x_i|^2) f'(x_i)]^{-1} \equiv z(1 + ze^{i\alpha})^{-2}. \tag{4}$$

---

[1] Received September 16, 1936; revised October 26, 1936.

[2] See J. E. Littlewood, *Proceedings of the London Mathematical Society* (2), vol 23 (1925), p. 482.

[3] See L. Bieberbach, *Sitzber. kgl. Akad. Berlin*, 1916, s. 940-955.

The following lemma is due to L. Bieberbach.[4]

LEMMA 1. *If* $f(z) = z + \sum_{2}^{\infty} a_n z^n$ *is holomorphic and univalent for* $|z| < 1$ *then* $|a_2| \leqq 2$ *where equality holds for, and only for, the function* $z(1 + ze^{i\alpha})^{-2}$, α *real.*

LEMMA 2. *Given* $f(z) = z + \sum_{2}^{\infty} a_n z^n$ *holomorphic for* $|z| < 1$, *and* x_0 *any fixed complex number such that* $|x_0| < 1$. *If we have identically, for some real constant* α,

$$(5) \qquad \left[f(x_0) - f\left(\frac{x_0 - z}{1 - \bar{x}_0 z}\right)\right] \cdot [(1 - |x_0|^2) \cdot f'(x_0)]^{-1} \equiv z(1 + e^{i\alpha} z)^{-2}$$

then $f(z)$ *must be identically equal to the function*

$$(6) \qquad f_0(z) \equiv \frac{z + (\bar{x}_0 - x_0 e^{2i\alpha})(x_0{}^2 e^{2i\alpha} - 1)^{-1} z^2}{\{1 - (\bar{x}_0 + e^{i\alpha})(x_0 e^{i\alpha} + 1)^{-1} z\}^2}.$$

Proof. Replacing z by $(x_0 - z)/(1 - \bar{x}_0 z)$ in (5) we obtain

$$(7) \qquad \frac{f(x_0) - f(z)}{(1 - |x_0|^2) f'(x_0)} \equiv (x_0 - z)(1 - \bar{x}_0 z) \cdot \{1 + x_0 e^{i\alpha} - (\bar{x}_0 + e^{i\alpha}) z\}^{-2}.$$

Since $f(0) = 0$, $f'(0) = 1$ we have

$$(8) \qquad f(x_0)[(1 - |x_0|^2) f'(x_0)]^{-1} = x_0(1 + x_0 e^{i\alpha})^{-2}$$

and

$$(9) \qquad [(1 - |x_0|^2) f'(x_0)]^{-1} = (1 - |x_0|^2)(1 - x_0 e^{i\alpha})(1 + x_0 e^{i\alpha})^{-3}.$$

Making the substitutions (8) and (9) in (7) we obtain for $f(z)$ the function $f_0(z)$ in (6) as required. We remark in passing that $f_0(z)$ is necessarily univalent for $|z| < 1$ since the right-hand side of the identity (5) is a function univalent in the unit circle, and since $(x_0 - z)/(1 - \bar{x}_0 z)$ is linear in z, mapping the unit circle onto itself.

LEMMA 3. *The function* $f_0(z)$ *of* (6) *above is such that it possesses a sequence of points* x_i *tending along a radius of the unit circle to the singularity of* $f_0(z)$ *on* $|z| = 1$ *for which* (4) *holds.*

This lemma may be easily verified by the reader.

THEOREM 1. *Let* $f(z) = z + \sum_{2}^{\infty} a_n z^n$ *be holomorphic and univalent for*

[4] See L. Bieberbach, *loc. cit.*

$|z| < 1$. *Let E denote the set of points on $|z| = 1$ at which $f(z)$ fails to be regular. Suppose that $f(z)$ has the property that there exists no sequence of points x_i of modulus inferior to unity, having a limit point in E for which (4) holds. Then there exists a $\mu > 0$, depending upon $f(z)$ but not upon r, and a constant $A(\mu)$ independent of r such that*

$$|f(re^{i\theta})| \leqq \frac{A(\mu) \cdot r}{(1-r)^{2-\mu}}, \qquad \mu < 2.$$

Proof. Let x be any point inside the unit circle. Let

$$\begin{aligned} F(z;x) &= \left[f(x) - f\left(\frac{x-z}{1-\bar{x}z}\right)\right] \cdot [(1-|x|^2)f'(x)]^{-1} \\ &= z + b_2(x)z^2 + \cdots. \end{aligned}$$

Since $f(z)$ is univalent and holomorphic for $|z| < 1$ so also is $F(z;x)$. On account of the hypothesis of our theorem relating to (4) we must have $f(z) \not\equiv f_0(z)$ by Lemma 3. Hence $F(z;x) \not\equiv z(1+ze^{i\alpha})^{-2}$ for any real value of α by Lemma 2. It follows from Lemma 1 that $|b_2(x)| < 2$. We may then write

$$\underset{|x|<1}{l.\,u.\,b.}\, |b_2(x)| = 2 - \mu(f), \qquad 0 \leqq \mu(f) \leqq 2.$$

We wish now to show that $\mu(f) > 0$. Suppose, if possible, $\mu(f) = 0$. Then there exists a sequence of points $\{x_i\}$ of modulus inferior to unity, converging to a point x_0, $|x_0| \leqq 1$, and a corresponding sequence of functions

$$F_i(z) \equiv F(z;x_i) = z + b_2(x_i)z^2 + \cdots$$

where $\lim\limits_{i\to\infty} |b_2(x_i)| = 2$. Since the functions $F(z;x_i)$ are univalent for $|z| < 1$ and therefore form a normal family in any domain interior to the unit circle, there exists an analytic function

$$F_0(z) = z + b_2{}^0 z^2 + \cdots$$

such that the functions $F(z;x_i)$ converge uniformly to $F_0(z)$ in any domain completely interior to the unit circle. Moreover,

$$|b_2{}^0| = \lim_{x_i \to x_0} |b_2(x_i)| = 2.$$

Since the convergence of the functions is uniform and $F_0(z)$ is not a constant,[5] $F_0(z)$ is univalent for $|z| < 1$. By Lemma 1

[5] See E. C. Titchmarsh, *On the Theory of Functions*, Oxford, 1932, page 200, 6.44.

$$F_0(z) \equiv z(1+ze^{i\alpha})^{-2}$$

for some real value α, since $|b_2^0| = 2$. There are now several cases to consider, depending upon whether $|x_0| < 1$ or $|x_0| = 1$. Suppose first $|x_0| < 1$. Then

$$\frac{f(x_0) - f\left(\dfrac{x_0 - z}{1 - \bar{x}_0 z}\right)}{(1 - |x_0|^2) f'(x_0)} \equiv \lim_{x_i \to x_0} F(z; x_i) \equiv F_0(z) \equiv z(1+ze^{i\alpha})^{-2}.$$

But this is impossible unless $f(z)$ be the function $f_0(z)$ of (6) which by hypothesis we have excluded (since it follows from (4) and Lemma 3 as we remarked above).

Secondly, suppose $|x_0| = 1$ and that $x_0 \subset E$. This case is clearly impossible by the hypothesis of our theorem. Thirdly, suppose $|x_0| = 1$ and x_0 is a point of regularity of $f(z)$. Since

$$\frac{x_i - z}{1 - \bar{x}_i z} = x_i + \frac{(|x_i|^2 - 1)z}{1 - \bar{x}_i z},$$

$$F_i(z) = \frac{z}{1 - \bar{x}_i z} + \sum_{n=2}^{\infty} \frac{f^{(n)}(x_i)}{n!\, f'(x_i)} \cdot (|x_i|^2 - 1)^{n-1} \left(\frac{z}{1 - \bar{x}_i z}\right)^n.$$

Since for every i both x_0 and x_i are points of regularity for $f(z)$ there exists a positive ρ independent of i such that $f(z)$ is regular in each circle of radius ρ having a centre at x_0 or x_i. Hence by Cauchy's inequality we have for some constant M depending only upon $f(z)$ and not upon n or i

$$\left| \frac{f^{(n)}(x_i)}{n!} \right| \leqq \frac{M}{\rho^n}.$$

$$\therefore \lim_{i \to \infty} \left| F_i(z) - \frac{z}{1 - \bar{x}_0 z} \right| \leqq \lim_{i \to \infty} (1 - |x_i|^2) P(z; x_i) = 0,$$

where $P(z; x_i)$ is a bounded quantity for all i. Hence in case x_0 is a point of regularity of $f(z)$ on $|z| = 1$ we have

$$F_0(z) \equiv \lim_{i \to \infty} F(z; x_i) \equiv \frac{z}{1 - \bar{x}_0 z}$$

which contradicts our assumption that $F_0(z) \equiv z(1+ze^{i\alpha})^{-2}$. Consequently we have shown under the assumptions of our theorem that $\mu(f) > 0$.

We now can write

$$|b_2(x)| = \left| \tfrac{1}{2} \frac{f''(x)}{f'(x)} (1 - |x|^2) - \bar{x} \right| \leqq 2 - \mu.$$

Let $x = re^{i\theta}$. On integrating we obtain

$$|\log f'(x) + \log(1-r^2)| \leqq \log\left(\frac{1+r}{1-r}\right)^{2-\mu}$$

$$|f'(re^{i\theta})| \leqq \frac{(1+r)^{1-\mu}}{(1-r)^{3-\mu}}$$

$$|f(re^{i\theta})| \leqq \int_0^r \frac{(1+r)^{1-\mu}}{(1-r)^{3-\mu}}\,dr < 2\int_0^r \frac{dr}{(1-r)^{3-\mu}}$$

If $\mu \neq 2$ then

$$|f(re^{i\theta})| \leqq \frac{2}{2-\mu}\left[\frac{1-(1-r)^{2-\mu}}{(1-r)^{2-\mu}}\right] < \frac{A(\mu)\cdot r}{(1-r)^{2-\mu}}.$$

If $\mu = 2$ then

$$|f(re^{i\theta})| \leqq \tfrac{1}{2}\log\left(\frac{1+r}{1-r}\right).$$

This completes the proof of our theorem.

THEOREM 2. *Let $f(z) = z + \sum_2^\infty a_n z^n$ satisfy the hypotheses of Theorem 1. Then there exist two positive constants δ and $A(\delta)$ depending upon $f(z)$, but not upon r, where $0 < \delta < 1$ such that*

$$\frac{1}{2\pi}\int_0^{2\pi} |f(re^{i\theta})|\,d\theta < \frac{A(\delta)\cdot r}{(1-r)^{1-\delta}}.$$

Proof. By Theorem 1 we can find a value of δ for which $0 < \delta < 1$ so that

$$|f(re^{i\theta})| < \frac{A(\delta)\cdot r}{(1-r)^{2-\delta}}$$

where $A(\delta)$ is a constant independent of r. If we define

$$F(\zeta) = \sqrt{f(\zeta^2)}, \qquad |\zeta| = \rho = r^{1/2},$$

then

$$|F(\zeta)| \leqq A(\delta)\cdot\rho(1-\rho^2)^{-1+\delta/2}.$$

Further, we have the following inequality due to J. E. Littlewood.[6]

$$\frac{1}{2\pi}\int_0^{2\pi} |f(re^{i\theta})|\,d\theta \leqq 2\int_0^\rho \max_{|\zeta|=\rho} |F(\zeta)|^2 \frac{d\rho}{\rho}.$$

$$\therefore \frac{1}{2\pi}\int_0^{2\pi} |f(re^{i\theta})|\,d\theta \leqq A(\delta)\int_0^\rho \frac{2\rho\,d\rho}{(1-\rho^2)^{2-\delta}}, \qquad \rho^2 = r$$

$$< \frac{A(\delta)\cdot r}{(1-r)^{1-\delta}}$$

where $A(\delta)$ is not necessarily the same constant in each place it appears above.

[6] See J. E. Littlewood, *loc. cit.*

14

THEOREM 3. *Let* $f(z) = z + \sum_2^\infty a_n z^n$ *satisfy the hypothesis of Theorem* 1. *Then there exists a* $\delta = \delta(f) > 0$, *depending upon* $f(z)$ *but not upon* n *such that*

$$a_n = O(n^{1-\delta}).$$

Proof. Since

$$| a_n | r^n \leqq \frac{1}{2\pi} \int_0^{2\pi} | f(re^{i\theta}) | \, d\theta < \frac{A(\delta) r}{(1-r)^{1-\delta}}$$

we obtain for $r = 1 - 1/n$, $n > 1$,

$$\begin{aligned} | a_n | &\leqq A(\delta) \left(1 - \frac{1}{n}\right)^{-n+1} \cdot n^{1-\delta} \\ &\leqq A(\delta) \left(1 + \frac{1}{n-1}\right)^{n-1} \cdot n^{1-\delta} \\ &< eA(\delta) \cdot n^{1-\delta}, \qquad n > 1. \end{aligned}$$

We now have immediately

COROLLARY. *If* $f(z)$ *satisfies the condition* (1) *it must then also possess the property* (4).

YALE UNIVERSITY.

INDEPENDENT SETS OF POSTULATES FOR ABELIAN GROUPS AND FIELDS IN TERMS OF THE INVERSE OPERATIONS.[1]

By DAVID G. RABINOW.

In a previous paper[2] I discussed an independent set of postulates for a group in terms of the inverse operation. In part I of the present paper I shall consider a simplified set of independent postulates for an Abelian group in terms of the inverse operation, while in part II independent postulates for a field in terms of the inverse operations will be developed.

For convenience let us here recapitulate the results of the previous paper referred to. The base is $(K, -)$ where K is a class of elements $a, b, c, \cdots$ and $-$ is a binary operation.

Postulate I. a in K and b in K imply $a - b$ in K.
Postulate II. $a - a = b - b$.
Definition I. $z = a - a$.
Definition II. $a' = z - a$.
Postulate III. $(a - b') - c' = a - (b - c')'$.
Postulate IV. $a'' = a$.
Postulate V. $a - b' = b - a'$.
Definition III. $a + b = a - b'$.

The results obtained were:

Result I. Any system $(K, -)$ which satisfies Postulates I through IV is a group with respect to $+$ as defined in Definition III.

Result II. Any system $(K, -)$ which satisfies Postulates I through V is an Abelian group with respect to the operation $+$.

Result III. An Abelian group is also given by I, II, III, V, and 11 where Theorem 11 is: $a - z = a$.

Result IV. Theorem 11 is deducible from I, II, III, and IV.

Result V. I, II, III, IV, and V are independent. Also I, II, III, V, and 11 are independent.

Hereafter we shall refer to these results as RI, RII, RIII, RIV, and RV respectively.

[1] Received June 16, 1936; revised October 5, 1936.

[2] "Independent set of postulates for a group in terms of the inverse operation," offered to the *Bulletin of the American Mathematical Society*.

PART I.

It is obvious from the above set of postulates that the definition of an Abelian group in terms of the inverse operation is much more complicated than the definition in terms of the direct operation. I shall now consider an independent set of postulates for an Abelian group in terms of the inverse operation which is as simple as the customary definition [3] in terms of the direct operation.

Let us consider the following set of postulates in connection with the base $(K, -)$ where K is a class of elements $a, b, c, \cdots$ and $-$ is a binary operation.

Postulate 1. a in K and b in K imply $a - b$ in K.

Postulate 2. $(a - b) - c = (a - c) - b$.

Postulate 3. $a - (a - b) = b$.

The following theorems are deducible from Postulates 1, 2 and 3.

THEOREM 1. *If* $a - b = a - c$, *then* $b = c$.

by hypothesis $a - b = a - c$

by Postulate 1 $a - (a - b) = a - (a - c)$

hence by Postulate 3 $b = c$.

THEOREM 2. $b - b = c - c$ *for every element.*

by Postulate 3 $b = a - (a - b)$

by Postulate 1 $b - b = [a - (a - b)] - b$

by Postulate 2 $\quad = (a - b) - (a - b)$

hence by Postulate 1 $(a - b) - (b - b) = (a - b) - [(a - b) - (a - b)$

by Postulate 3 $\quad = a - b \qquad \text{(I)}$

now by Postulate 2 $(a - a) - b = (a - b) - a$

hence $(a - b) - [(a - a) - b] = (a - b) - [(a - b) - a$

by Postulate 3 $\quad = a \qquad \text{(II)}$

hence by (I) $[(a - b) - (b - b)] - [(a - a) - b] = a$

by Postulate 2 $\{(a - b) - [(a - a) - b]\} - (b - b) = a$

hence by (II) $a - (b - b) = a$

[3] For bibliographical references the reader may consult: E. V. Huntington, "Note on the definition of abstract groups and fields by sets of independent postulates," *Transactions of the American Mathematical Society*, vol. 6 (1905), pp. 181-193; and R. Garver, "Postulates for special types of groups," *Bulletin of the American Mathematical Society*, vol. 42 (1936), pp. 125-129.

however b is any element of K, therefore

$a - (c - c) = a$, where c is any element of K

hence $a - (b - b) = a - (c - c)$

hence by Theorem 1 $b - b = c - c$.

DEFINITION 1. $z = a - a$.

DEFINITION 2. $a' = z - a$.

THEOREM 3. $a'' = a$.

by Postulate 3 $b - (b - a) = a$ for all elements b

by Definition 1 $z - (z - a) = a$

by Definition 2 $a'' = a$.

THEOREM 4. $a - b = (b - a)'$.

by Postulate 3 $b - (b - a) = a$

by Postulate 1 $[b - (b - a)] - b = a - b$

by Postulate 2 $(b - b) - (b - a) = a - b$

by Definition 1 $z - (b - a) = a - b$

by Definition 2 $(b - a)' = a - b$.

THEOREM 5. $a - b = b' - a'$.

by Definition 2 $b' - a' = (z - b) - (z - a)$

by Postulate 2 $= [z - (z - a)] - b$

by Theorem 3 $= a - b$.

THEOREM 6. $(a - b') - c' = a - (b - c')'$.

by Theorem 4 $(b - c')' = c' - b$

by Postulate 1 $a - (b - c')' = a - (c' - b)$

by Theorem 4 $= [(c' - b) - a]'$

by Postulate 2 $= [(c' - a) - b]'$

by Theorem 4 $= b - (c' - a)$

by Theorem 4 $= b - (a - c')'$

by Theorem 5 $= (a - c') - b'$

by Postulate 2 $= (a - b') - c'$.

THEOREM 7. $a - b' = b - a'$.

by Theorem 6 $(a - b') - (b - a') = a - [b - (b - a')]'$

by Postulate 3 $= a - a''$

by Theorem 3 $= a - a$

by Definition 1 $= z$

now by Definition 1 $(a - b') - (a - b') = z$

hence $(a - b') - (a - b') = (a - b') - (b - a')$

by Theorem 1 $a - b' = b - a'$.

DEFINITION 3. $a + b = a - b'$.

However by RII we know that Postulate 1, Theorems 2, 3, 6, and 7 are the postulates for an Abelian group with respect to the operation $+$ defined in Definition 3. Hence Postulates 1, 2, and 3 define an Abelian group with respect to $+$. It remains to show that Postulates 1, 2, and 3 are deducible from RII; that is, we have shown that any system $(K, -)$ which satisfies Postulates 1, 2, and 3 is an Abelian group and we must show conversely, that any Abelian group satisfies Postulates 1, 2, and 3.

THEOREM 8. a in K and b in K imply $a - b$ in K.

by Postulate I this is true.

THEOREM 9. $(a - b) - c = (a - c) - b$.

by Postulate III $(a - b) - c = a - (b' - c)'$

by Postulate V $= a - (c' - b)'$

by Postulate III $= (a - c) - b$.

THEOREM 10. $a - (a - b) = b$.

by Theorem 8 $a - (a - b)$ is in K and hence $a - (a - b) = X$

by Theorem 8 $[a - (a - b)] - (a - b)' = X - (a - b)'$

by Postulate III $a - [(a - b)' - (a - b)']' = X - (a - b)'$

by Definitions I and II $a - z = X - (a - b)'$

by RIV and Postulate III $a = (X - a') - b$

by Theorem 8 $a - b' = [(X - a') - b] - b'$

hence by Postulate III, Definitions I and II, and RIV

$$a - b' = X - a'$$

by Postulate V $b - a' = X - a'$

by Theorem 8 $(b - a') - a = (X - a') - a$

by Postulate III, Definitions I and II, and RIV

$$b = X.$$

The postulates are examined for independence by the usual method of exhibiting examples of systems $(K, -)$ which fail to satisfy the correspondingly numbered postulate but satisfy the remaining postulates.

Example 1) K is the class of all prime positive integers with $a - b = a/b$.

Example 2) K is the class of all positive integers with $a - b = b$.

Example 3) K is the class of all positive integers with $a - b = a + b$.

PART II.

Let us consider the following set of postulates in connection with the base $(K, -, \circ)$ where K is a class of elements $a, b, c, \cdots$ and $-, \circ$ are binary operations.

Postulate E. K contains at least two distinct elements.

Postulate 11. a in K and b in K imply $a - b$ in K.

Postulate 12. $a - a = b - b$.

Definition 1. $z = a - a$.

Definition 2. $a' = z - a$.

Postulate 13. $(a - b') - c' = a - (b - c')'$.

Postulate 14. $a'' = a$.

Postulate 15. $a - b' = b - a'$.

Definition 3. $a + b = a - b'$.

Postulate 16. a in K and b in K and $b \neq z$ imply $a \circ b$ in K.

Postulate 17. $a \circ a = b \circ b$ if $a \circ a$ and $b \circ b$ are in K.

Definition 4. $U = a \circ a$ if $a \circ a$ is in K.

Definition 5. $a^* = U \circ a$ if $U \circ a$ is in K.

Postulate 18. 1) If U exists then $U \neq z$.

2) If a^* exists then $a^* \neq z$ (provided $a \neq z$).

Postulate 19. If $a, b, c, b^*, c^*, a \circ b^*, (b \circ c^*)^*, (a \circ b^*) \circ c^*$ and $a \circ (b \circ c^*)^*$ are in K, then $(a \circ b^*) \circ c^* = a \circ (b \circ c^*)^*$.

Postulate 20. If a, a^* and a^{**} are in K, then $a^{**} = a$.

Postulate 21. If $a, b, a^*, b^*, a \circ b^*$ and $b \circ a^*$ are in K, then $a \circ b^* = b \circ a^*$.

Postulate 22. 1) If $a, b, c, b - c', b^*, c^*, (b - c')^*, a \circ b^*, a \circ c^*, a \circ (b - c')^*$ and $a \circ b^* - (a \circ c^*)'$ are in K, then

$$a \circ (b - c')^* = a \circ b^* - (a \circ c^*)'.$$

2) If $a, b, c, b - c', a^*, b \circ a^*, (c \circ a^*)', (b - c') \circ a^*$ and $b \circ a^* - (c \circ a^*)'$ are in K, then

$$(b - c') \circ a^* = b \circ a^* - (c \circ a^*)'.$$

Definition 6. 1) $a \cdot b = a \circ b^*$ if $b \neq z$.

2) $a \cdot z = z$.

Note 1. Postulates 15, 20 and 22. 1 are redundant. They have been retained to keep the list of postulates symmetrical.

LEMMA. *The element U must exist.*

For by Postulate E and Definition 1 there exists an element distinct from z. Let this element be a. Then by Postulate 16 $a \circ a$ exists, and hence by Definition 4 the element U exists.

THEOREM 23. *$(K, -)$ forms an Abelian group with respect to the operation $+$ defined in Definition 3. This follows from* RII.

THEOREM 24. *a in K and b in K imply $a \cdot b$ in K.*

Case I. $b = z$.

by Definition 6.2	$a \cdot z = z$ for all a.

Case II. $b \neq z$.

by hypothesis	$b \neq z$
by Postulate 18.2	$b^* \neq z$
by Postulate 16	b^* is in K
by Postulate 16	aob^* is in K
by Definition 6.1	$a \cdot b$ in K.

THEOREM 25. *If $a - b = a - c$ then $b = c$.*

by hypothesis	$a - b = a - c$
by Definition 2	$(a - b)' = (a - c)'$
by Postulate 11	$a' - (a - b)' = a' - (a - c)'$
by Postulate 13	$(a' - a') - b = (a' - a') - c$
by Definition 1	$z - b = z - c$
by Postulate 14	$b = c$.

THEOREM 26. $z \cdot a = z$.

Case I. $a = z$.

by Definition 6.2	$a \cdot z = z$ for all a
hence	$z \cdot z = z$.

Case II. $a \neq z$.

by Definition 6.1	$a \cdot b = aob^*$ if $b \neq z$
hence	$z \cdot a = zoa^*$
by Theorem 23	$= zoa^* - z$
now by Definitions 1 and 2	$zoa^* = (z - z)oa^* = (z - z')oa^*$
by Postulate 22.2	$= zoa^* - (zoa^*)'$
hence	$zoa^* - z = zoa^* - (zoa^*)'$
by Theorem 25	$(zoa^*)' = z$

hence by Postulate 14 and Definition 2 $zoa^* = z$.

THEOREM 27. $a \cdot b = b \cdot a$.

Case I. $a = b = z$.

by Theorem 26	$z \cdot z = z \cdot z = z$

Case II. $b = z$, $a \neq z$.

by Definition 6.2 $\quad a \cdot z = z$

by Theorem 26 $\quad z \cdot a = z.$

Case III. $a \neq z$, $b \neq z$.

by Postulate 18.2 $\quad a^* \neq z$, $b^* \neq z$

by Postulate 16 $\quad aob^*$ and boa^* are in K

by Postulate 21 $\quad aob^* = boa^*$

by Definition 6.1 $\quad a \cdot b = b \cdot a.$

THEOREM 28. $U = U^*$.

by Postulate 18.1 $\quad U \neq z$

by Definition 5 $\quad U^* = UoU$

by Definition 4 $\quad = U.$

THEOREM 29. $boU = b$.

Case I. $b = z$.

Since by Postulate 18.1 $U \neq z$ then boU is in K by Postulate 16.

by Theorem 28 $\quad boU = boU^*$

hence $\quad zoU = zoU^*$

by Theorem 26 $\quad = z.$

Case II. $b \neq z$.

by Theorem 28 $\quad boU = boU^*$

by Definition 4 $\quad = bo(b^*ob^*)^* \quad$ (A)

now since by hypothesis $b \neq z$ hence by Postulate 18.2 $b^* \neq z$ and $b^{**} \neq z$

by Postulate 19, (A) $\quad boU = (bob^{**})ob^*$

by Postulate 20 $\quad = (bob)ob^*$

$= Uob^* = b^{**} = b$ by Definitions 4 and 5 and Postulate 20.

THEOREM 30. $a \neq z$ and $b \neq z$ *imply* $aob \neq z$.

by Postulate 20 $\quad aob = aob^{**}$

now suppose $aob = z$ then

by Theorem 26 $\quad (aob^{**})ob^* = zob^* = z$

$= ao(b^*ob^*)^* = aoU^* = aoU = a$ by Postulate 19, Definition 4, and Theorem 29.

hence $\quad a = z$

but this contradicts the hypothesis that $a \neq z$,

hence $\quad aob \neq z.$

THEOREM 31. $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.

Case I. $z = c$, $a \neq z$, $b \neq z$.

by Theorem 24	ab is in K
by Definition 6.2	$(a \cdot b) \cdot z = z$
by Definition 6.2	$b \cdot z = z$
by Definition 6.2	$a \cdot z = z$.

Case II. $b = z$.

by Definition 6.2	$az = z$
by Theorem 26	$z \cdot c = z$
by Theorem 26	$z \cdot c = z$
by Definition 6.2	$a \cdot z = z$.

Case III. $a = z$.

by Theorem 26	$z \cdot b = z$
by Theorem 26	$z \cdot c = z$
by Theorem 24	$b \cdot c$ is in K
by Theorem 26	$z \cdot (b \cdot c) = z$.

Case IV. $a = b = z$ and other combinations follow from Definition 6.2 and Theorem 26.

Case V. $a \neq z$, $b \neq z$, $c \neq z$.

by hypothesis $a \neq z$, $b \neq z$, $c \neq z$

hence by Postulate 18.2 and Theorem 30 $a^* \neq z$, $b^* \neq z$, $c^* \neq z$, $boc^* \neq z$

by Postulate 19	$ao(boc^*)^* = (aob^*)oc^*$
by Definition 6.1	$a \cdot (b \cdot c) = (a \cdot b) \cdot c$.

THEOREM 32.1. *There exists an element* $u \neq z$ *such that* $a \cdot u = a$.

by Theorem 29	$aoU = a$ for all a
by Theorem 28	$aoU = aoU^*$
by Postulates 18.1 and 18.2	$U^* \neq z$
by Definition 6.1	$a \cdot U = a$ for all a.

THEOREM 32.2. $u \cdot a = a$ *where* u *is the* u *of Theorem* 32.1.

Case I. $a = z$.

by Definition 6.2 $U \cdot z = z$.

Case II. $a \neq z$.

by Definition 6.1 $U \cdot a = Uoa^*$

$= a^{**} = a$ by Definition 5 and Postulate 20.

THEOREM 33. *If $a \neq z$ then there exists an element x such that $ax = b$.*

Case I. $b = z$.

by Definition 6.2 $\quad a \cdot z = z$.

Case II. $b \neq z$.

by Postulate 20 $\quad b = Uob^*$

by Definition 4 $\quad = (aoa)ob^*$

by Postulate 19 $\quad = ao(a^*ob^*)^*$

by hypothesis and Theorem 30 $\quad (a^*ob^*)^* \neq z$

hence by Definition 6.1 $\quad b = a \cdot x$ where $x = a^*ob^*$.

THEOREM 34. *If $a \neq z$ then there exists an element y such that $ya = b$.*

Case I. $b = z$.

by Theorem 26 $\quad z \cdot a = z$.

Case II. $b \neq z$.

by Theorem 29 $\quad b = boU$

$\quad = bo(a^*oa^*)^*$ by Definition 4 and Theorem 28.

by hypothesis $\quad a \neq z$, hence by Postulate 18.2 $a^* \neq z$, then

by Postulate 19 $\quad boU = (boa)oa^*$

by Definition 6.1 $\quad b = y \cdot a$ where $y = boa \quad$ (A).

THEOREM 35. *If $y_1 \cdot a = b$ and $y_2 \cdot a = b$, then $y_1 = y_2$ if $a \neq z$.*

Case I. $b = z$.

by hypothesis $\quad y_1 \cdot a = y_2 \cdot a$. Now if $y_1 \neq z$ or $y_2 \neq z$ and by hypothesis $a \neq z$, then by Theorem 30 either $y_1oa^* \neq z$ or $y_2oa^* \neq z$.

hence by Definition 6.1 either $y_1 \cdot a \neq z$ or $y_2 \cdot a \neq z$. Since neither of these is possible, then $y_1 = y_2 = z$.

Case II. $b \neq z$.

by Theorem 33 there exists an element x such that $a \cdot x = U$

hence $\quad (y_1 \cdot a) \cdot x = (y_2 \cdot a) \cdot x$

by Theorem 31 $\quad y_1 \cdot (a \cdot x) = y_2 \cdot (a \cdot x)$

hence $\quad y_1 \cdot U = y_2 \cdot U$

by Theorem 32.1 $\quad y_1 = y_2$.

THEOREM 36. *The element y of Theorem 34 is uniquely determined by a and b.*

by (A) of Theorem 34 $\quad y = boa$.

Hence the operation o is the inverse of the operation $\cdot$.

THEOREM 37. $(a+b)\cdot c = a\cdot c + b\cdot c$.

Case I. $c = z$.

by Definition 6.2 $\quad (a+b)\cdot c = z,\ a\cdot z = z,\ b\cdot z = z$

by Definitions 1, 2, and 3 $\quad z + z = z$.

Case II. $c \neq z$.

by Postulate 22.2 $\quad (a - b')oc^* = aoc^* - (boc^*)'$

by Definition 6.1 $\quad (a - b')\cdot c = a\cdot c - (b\cdot c)'$

by Definition 3 $\quad (a+b)\cdot c = a\cdot c + b\cdot c$.

THEOREM 38. $a\cdot(b+c) = a\cdot b + a\cdot c$.

by Theorem 37 $\quad (b+c)\cdot a = b\cdot a + c\cdot a$

by Theorem 27 $\quad a\cdot(b+c) = a\cdot b + a\cdot c$.

THEOREM 39. $(aob^*)' = a'ob^*$ *if* $b \neq z$.

by Theorem 26 $\quad zob^* = z$

by Definitions 1 and 2 $\quad (a' - a')ob^* = z$

by Postulate 22.2 $\quad a'ob^* - (aob^*)' = z$

by Definition 1 $\quad a'ob^* - a'ob^* = z$

hence $\quad a'ob^* - a'ob^* = a'ob^* - (aob^*)'$

by Theorem 25 $\quad a'ob^* = (aob^*)'$.

THEOREM 40. $a - b = (b - a)'$.

by Postulate 13 $\quad (a-b) - (b-a)' = a - [b' - (b-a)']'$

by Postulate 13 $\quad = a - [(b' - b') - a]'$

by Definition 1 $\quad = a - (z - a)'$

$= a - a'' = a - a = z$ by Definitions 1 and 2 and Postulate 14

by Definition 1 $\quad (a-b) - (a-b) = z$

hence $\quad (a-b) - (a-b) = (a-b) - (b-a)'$

by Theorem 25 $\quad a - b = (b-a)'$.

THEOREM 41. *If* $b - a = c - a$, *then* $b = c$.

by hypothesis $\quad b - a = c - a$

by Postulate 11 $\quad (b-a) - a' = (c-a) - a'$

by Postulate 13 $\quad b - (a' - a')' = c - (a' - a')'$

by Definitions 1 and 2 $\quad b - z = c - z$

by RIV $\quad b = c$.

THEOREM 42. $a - b' = b - a'$.

Case I. $a = b = z$.

by Definition 1 $\quad z - z = z - z = z$.

Case II. $a = z,\ b \neq z$ (also $b = z,\ a \neq z$).

by Definition 2 and Postulate 14 $\quad a - b' = z - b' = b'' = b$

by Definition 2 and RIV $\quad b - a' = b - z = b$.

Case III. $a \neq z,\ b \neq z,\ a - b' = z$.

by hypothesis $\quad a - b' = z$

by Definition 1 $\quad z = b' - b'$

hence $\quad a - b' = b' - b'$

by Theorem 41 $\quad a = b'$

by Postulate 14 $\quad a' = b$

by Definition 1 $\quad b - a' = z$.

Case IV. $a \neq z,\ b \neq z,\ a - b' \neq z$.

by Postulate 20 $\quad a - b' = U\text{o}(a - b')^*$

by Postulate 22. 1 $\quad = U\text{o}a^* - (U\text{o}b^*)'$

by Theorem 39 $\quad = U\text{o}a^* - U'\text{o}b^*$

by Theorem 39 $\quad = (U'\text{o}a^*)' - U'\text{o}b^*$

by Theorem 40 $\quad = [U'\text{o}b^* - (U'\text{o}a^*)']'$

by Postulate 22. 1 $\quad = [U'\text{o}(b - a')^*]'$

$\quad = U\text{o}(b - a')^*$ by Theorem 39 and Definition 5

by Postulate 20 $\quad = b - a'$.

THEOREM 43. $a^{**} = a$.

Case I. $a = z$. For a^{**} to have meaning in this case, a^* must be in K and a^{**} must be in K.

by Postulate 21 $\quad U\text{o}a^* = a\text{o}U^*$

hence $\quad U\text{o}z^* = z\text{o}U^*$

by Theorem 26 $\quad = z$.

Case II. $a \neq z$.

by Postulate 21 $\quad a\text{o}a^{**} = a^*\text{o}a^* \quad$ (A)

by Definition 4 $\quad = U$

by Postulate 22. 2 $\quad (a - a^{**})\text{o}a^{**} = a\text{o}a^{**} - (a^{**\prime}\text{o}a^{**})'$

by Theorem 39 $\quad = a\text{o}a^{**} - a^{**}\text{o}a^{**}$

by Definition 4 and (A) $\quad = U - U$

by Definition 1 $\quad = z \quad$ (B)

by Theorem 26 and (B) $[(a - a^{**})oa^{**}]oa^* = zoa^* = z$
by Postulate 19 $(a - a^{**})o(a^*oa^*)^* = z$
by Definition 4 $(a - a^{**})oU^* = z$ (E)
now let us suppose $a - a^{**} \neq z$
by Postulate 21, (E) $(a - a^{**})oU^* = Uo(a - a^{**})^* = z$ (C)
if $a - a^{**} \neq z$, then
by Postulate 18.2 $(a - a^{**})^* \neq z$ and $Uo(a - a^{**})^* \neq z$
but by (C) we see $Uo(a - a^{**})^* = z$
hence the assumption that $a - a^{**} \neq z$ leads to a contradiction and hence

$$a - a^{**} = z$$

by Theorem 25 $a = a^{**}$.

THEOREM 44. $ao(b - c')^* = aob^* - (aoc^*)'$.

Case I. $a = z,\ b \neq z,\ c \neq z,\ b - c' \neq z.$
by Theorem 26 $zo(b - c')^* = zob^* = zoc^* = z.$

Case II. $b - c' = z, b \neq z, c \neq z, a \neq z.$ For this case to have meaning $ao(b - c')^*$ must be in K.
by Postulate 22.2 $(b - c')oa^* = boa^* - (coa^*)'$
by Postulate 21 $ao(b - c')^* = aob^* - (aoc^*)'$.

Case III. $b - c' = z, b = z, c \neq z, a \neq z.$ The proof of this case is the same as Case II.

Case IV. $b - c' = z, b = z, c = z, a \neq z.$ The proof of this case is the same as Case II.

Case V. $b - c' = z, b \neq z, c \neq z, a = z.$ The proof of this case is the same as Case II.

Case VI. $a \neq z, b \neq z, c \neq z, b - c' \neq z.$
by Postulate 22.2 $(b - c')oa^* = boa^* - (coa^*)'$
by Postulate 21 $ao(b - c')^* = aob^* - (aoc^*)'$.

However Theorems 23, 24, 27, 31, 32.1, 32.2, 33, 34, 37 and 38 are the postulates for a field in terms of the operations $+, \cdot$.[4] Hence any system $(K, -, o)$ which satisfies Postulates 11 through 22.2 is a field in terms of the operations $+, \cdot$.

The postulates are examined for independence by exhibiting examples of systems $(K, -, o)$ which fail to satisfy the correspondingly numbered postulates but satisfy the remaining postulates.

[4] See reference to Huntington's paper in footnote 3.

Example 11. K is the class of three numbers 0, 1, 2 with $a - b$ and aob satisfying the following multiplication tables. The elements w and u which appear in the tables are elements not in K.

$a - b$	0	1	2
0	0	2	1
1	1	0	w
2	2	w	0

aob	0	1	2
0	u	0	0
1	u	1	2
2	u	2	1.

Example 12. K is the class of all positive rational numbers including zero. $a - b = a + b$ with $z = 0$. $aob = a/b$.

Example 13. K is the class of all positive rational numbers including zero. $a - b = a - b$. $aob = a/b$.

Example 14. K is the class of three numbers 0, 1, 2 with $a - b$ and aob satisfying the following multiplication tables. The element u which appears in the table for o is an element not in K.

$a - b$	0	1	2
0	0	0	0
1	0	0	1
2	0	1	0

aob	0	1	2
0	u	0	0
1	u	1	2
2	u	2	1.

Example 16. K is the class of all integers, positive, negative, and zero. $a - b = a - b$. $aob = a/b$.

Example 17. K is the class of all rational numbers, positive, negative, and zero. $a - b = a - b$. $aob = ab$ with $U = 1$.

Example 18. 1. K is the class of all rational numbers, positive, negative, and zero. $a - b = a - b$. $aob = (a - a)/b$, that is equals zero. Here $U = 0$.

Example 18. 2. K is the class of all rational numbers, positive, negative, and zero. $a - b = a - b$. $aob = a/b$ except $1/a = 0$.

Example 19. K is the class of hypercomplex numbers of the form $\pi 1 + \omega i + \rho j$ where π, ω, ρ are rational numbers, positive, negative, and zero. $a - b = a - b$.

$$aob = \frac{1}{\pi_2{}^2 + \rho_2{}^2 + (\omega_2 - \rho_2)^2} (\pi_1 1 + \omega_1 i + \rho_1 j)(\pi_2 1 + \omega_2 i + \rho_2 j),$$

where the product of the coefficients shall be the ordinary product of rational numbers and the "units" shall follow the table,

	1	i	j
1	1	$-i$	$-j$
i	i	1	-1
j	j	-1	2.

Example 21. K is the class of quaternions. $a - b = a - b$.

$$aob = \frac{1}{\pi_2^2 + \omega_2^2 + \rho_2^2 + \delta_2^2} (\pi_1 1 + \omega_1 i + \rho_1 j + \delta_1 k)(\pi_2 1 + \omega_2 i + \rho_2 j + \delta_2 k)$$

where the product of the coefficients shall be the ordinary product of rational numbers, while the "units" shall obey the table

	1	i	j	k
1	1	$-i$	$-j$	$-k$
i	i	1	$-k$	j
j	j	k	1	$-i$
k	k	$-j$	i	1.

Example 22. 2. K is the class of two numbers 0, 1. $a - b$ and aob satisfy the following tables. The element u in the table for o is an element not in K.

$a - b$	0	1
0	0	1
1	1	0

aob	0	1
0	u	1
1	u	1.

Example E. K is the class consisting of 0 only with $0 - 0 = 0$ and $0o0$ undefined.

HARVARD UNIVERSITY.

... LIBRARY. Acc. No. ... Book No. ...

ASTRONOMICAL CONSEQUENCES OF THE RELATIVISTIC TWO-BODY PROBLEM.*

By Tullio Levi-Civita.

1. Mechanical laws, according to Einstein's theory, are much more complicated in conception than under the assumptions of Newton. However the motion of celestial bodies under ordinary circumstances differ so little from their Newtonian representation, that, for astronomical purposes, relativistic effects may be conveniently treated as first-order perturbations.

A good amount of work in this direction was done, shortly after the appearance of general relativity, with deep insight and high competence by the late Professor De Sitter.[1]

The simple case of two bodies of *comparable* masses lies beyond De Sitter's developements, which were chiefly directed towards the inclusion of perturbations arising from relativity in the standard equations concerning planets and satellites of our solar system, where one of the masses predominates.

I have recently taken up the question,[2] paying due attention to the case of comparable masses. For the usual two-body problem, which in the traditional hierarchy comes immediately after Einstein's one-centre problem, the equations of motion are certainly integrable if one treats relativistic effects as first-order perturbations.

We intend to say a few words about deduction and illustration of two inequalities which are already apparent or, at least, may shortly appear in the observable field.

2. Let us start from the explicit form of the two Lagrangian functions

$$L_0 \text{ and } L_1$$

which define the *absolute* motion of the centres of mass, P_0 and P_1, of two celestial bodies; e. g., a double star.

I suppose that everything has already been reduced to ordinary space, $x_h{}^i$

* A paper delivered at the Tercentenary Conference of Arts and Sciences at Harvard University, September 4, 1936. Received by the Editors January 18, 1937.

[1] *Monthly Notices, Royal Astronomical Society*, vol. 77 (1916), pp. 155-184 (Second paper).

[2] "The relativistic problem of several bodies," *American Journal of Mathematics*, vol. 59 (1937), pp. 9-22.

$(i = 1, 2, 3)$ being Cartesian coördinates of P_h $(h = 0, 1)$ with reference to some fixed or Galilean frame, while the independent variable is $x^0 = ct$ (t usual time and c velocity of light).

The Lagrangian equations, furnished by the L_h $(h = 0, 1)$, define the components

$$\ddot{x}_h{}^i \qquad (h = 0, 1)$$

of the absolute accelerations of the two points P_h as functions of positions and velocities. At this stage, all textbooks introduce relative coördinates

$$x^i = x_1{}^i - x_0{}^i \qquad (i = 1, 2, 3)$$

and corresponding relative accelerations

$$\ddot{x}^i = \ddot{x}_1{}^i - \ddot{x}_0{}^i \qquad (i = 1, 2, 3)$$

simply by subtraction. Then, on account of the fact that all but the Newtonian terms are of the second order, we are allowed to use Keplerian values, and especially to employ the classical integrals of energy and areas.

If the simple but rather tedious developments are performed with some insight into the matter, we recognize that relative motion may also be brought under the Lagrangian scheme. The corresponding function L is

$$L = N + \Pi, \tag{1}$$

where, to within a constant factor, N designates the usual Newtonian term and Π the additional relativistic contribution. More precisely, if m_0, m_1 are the masses of the two bodies and $r = \overline{P_0P_1}$ is their mutual distance, then putting

$$m = m_0 + m_1 \tag{2}$$

and

$$\beta_i = \frac{dx^i}{dx^0} = \dot{x}^i, \qquad \beta^2 = \sum_1^3{}_i \beta_i{}^2, \qquad \gamma = \frac{fm}{c^2}\,\frac{1}{r}, \tag{3}$$

we obviously have

$$N = \tfrac{1}{2}\beta^2 + \gamma. \tag{4}$$

In the expression of Π we shall denote by $\mathfrak{e}$ the difference

$$\tfrac{1}{2}\beta^2 - \gamma$$

which, in Newtonian approximation, is nothing but the constant of energy divided by c^2, so that, up to terms of second order, the numerical value of $\mathfrak{e}$ behaves like a constant. On the other hand, it is not the same to apply the Lagrangian operator

$$\frac{d}{dx^0}\,\frac{\partial}{\partial\beta_i}-\frac{\partial}{\partial x^i}$$

to the constant $\mathfrak{e}$, which gives zero, as to the binomial $\frac{1}{2}\beta^2-\gamma$, which gives $\dot{\beta}_i+\partial\gamma/\partial x^i$, or, up to the first order, $2\partial\gamma/\partial x^i$. Therefore, if we include $\mathfrak{e}$ among the arguments by means of which the explicit expression of Π is built, it is necessary to state whether $\mathfrak{e}$ is to be treated, in performing Lagrangian operations, as a genuine constant or as the difference $\frac{1}{2}\beta^2-\gamma$, which, as far as first approximation is concerned, has the same numerical value.

Attributing to $\mathfrak{e}$ at any moment the rôle of a simple constant, I have obtained

$$\Pi=(2-\tfrac{1}{2}\mathfrak{p})\beta^2\gamma-(1+\tfrac{1}{2}\mathfrak{p})\gamma^2+2(-1+2\mathfrak{p})\mathfrak{e}\gamma+\tfrac{1}{2}\mathfrak{p}\frac{1}{\mathfrak{a}^2}\gamma^3, \tag{5}$$

where

$$\mathfrak{p}=\frac{m_0m_1}{m^2} \tag{6}$$

and $1/\mathfrak{a}$ is a further dimensionless constant, connected, within terms of higher order, with the double areal velocity C by the relation

$$\frac{1}{\mathfrak{a}}=\frac{C}{c.\,fm/c^2}\,. \tag{7}$$

Accordingly, in ordinary planetary pairs and double stars, $\mathfrak{a}\sim\beta$, i. e., $\mathfrak{a}$ has the same order of magnitude as β. In particular, for circular motion, $\mathfrak{a}$ is the constant value of β. This is verified at once by means of the elementary relations

$$\beta^2=\frac{fm}{c^2}\,\frac{1}{a},\qquad \frac{C}{c}=a\beta,$$

in which a means the radius or, more generally, the major semi-axis of the orbit.

Solving for fm/c^2 and C/c, and substituting in (7), we get simply

$$\frac{1}{\mathfrak{a}}=\frac{1}{\beta}\,.$$

At any rate, from $\mathfrak{a}\sim\beta$ and $\gamma\sim\beta^2$, it follows that

$$\frac{1}{\mathfrak{a}^2}\gamma^3\sim\gamma^2,$$

showing that the last term of Π is, like the others, of the second order.

Squaring (7) and remembering the classical relation

$$C^2 = fma(1 - e^2) \qquad (e = \text{eccentricity}),$$

we get

$$\frac{1}{\mathfrak{a}^2}\,\frac{fm}{c^2} = a(1 - e^2), \tag{7'}$$

which will be used later on.

3. The expression (1) of L has, in view of (3) and (5), the form

$$L = \tfrac{1}{2}\psi\beta^2 + \phi, \tag{8}$$

where both ψ and ϕ depend exclusively on the mutual distance r, since

$$\psi = 1 + (4 - \mathfrak{p})\gamma, \tag{9}$$

$$\phi = \gamma - (1 + \tfrac{1}{2}\mathfrak{p})\gamma^2 + 2(-1 + 2\mathfrak{p})\mathfrak{e}\gamma + \tfrac{1}{2}\mathfrak{p}\frac{1}{\mathfrak{a}^2}\gamma^3. \tag{10}$$

The motion defined by the Lagrangian function L admits the integral

$$\tfrac{1}{2}\psi\beta^2 - \phi = \mathfrak{e}^*, \tag{11}$$

where the constant $\mathfrak{e}^*$ may differ from $\mathfrak{e}$ only in second order terms.

Now an equivalence theorem in analytical dynamics[3] states that the Lagrangian function

$$L_1 = \tfrac{1}{2}\beta^2 + \psi(\phi + \mathfrak{e}^*), \tag{12}$$

for which the motion admits the integral

$$\tfrac{1}{2}\beta^2 - \psi(\phi + \mathfrak{e}^*) = \text{const.}, \tag{13}$$

gives rise, for the value 0 of the constant in the second member, to a family of trajectories identical with those defined by (8) and the integral (11).

Therefore, as far as trajectories are concerned, our task is reduced to characterize those belonging to the Lagrangian function (12). The latter corresponds to the motion of a free particle in ordinary space under a conservative (even central) force having the force-function

$$\Phi = \psi(\phi + \mathfrak{e}^*).$$

Omitting the additive constant $\mathfrak{e}^*$ and all terms of order higher than two, and writing correspondingly $\mathfrak{e}$ instead of $\mathfrak{e}^*$ in all second order terms, we have, by (9) and (10),

[3] Cf., e. g., Levi-Civita and Amaldi, *Lezioni di meccanica razionale*, vol. II_2 (Bologna, Zanichelli, 1927), pp. 514-515.

$$\Phi = \gamma + (2 + 3\mathfrak{p})\mathfrak{e}\gamma + 3(1 - \tfrac{1}{2}\mathfrak{p})\gamma^2 + \frac{\mathfrak{p}}{\mathfrak{a}^2}\gamma^3.$$

Putting for a moment

$$m^* = m\{1 + (2 + 3\mathfrak{p})\mathfrak{e}\} \quad \text{and} \quad \gamma^* = \frac{fm^*}{c^2}\,\frac{1}{r} = \gamma + (2 + 3\mathfrak{p})\mathfrak{e}\gamma,$$

we may write simply γ^* instead of γ in the second and third terms of Φ. Here, however, it is indifferent, up to the second order, whether one employs γ or γ^*. Hence, omitting asterisks, we may consider the trajectories of a central force with the potential function

$$\text{(I)} \qquad \gamma + 3(1 - \tfrac{1}{2}\mathfrak{p})\gamma^2 + \tfrac{1}{2}\frac{\mathfrak{p}}{\mathfrak{a}^2}\gamma^3,$$

where, from (2), (3), (6) and (7),

$$\text{(II)} \quad m = m_0 + m_1, \qquad \gamma = \frac{fm}{c^2}\,\frac{1}{r}, \qquad \mathfrak{p} = \frac{m_0 m_1}{m^2}, \qquad \frac{1}{\mathfrak{a}} = \frac{C}{cfm/c^2}.$$

Note that, in these formulae (owing to the described policy of first introducing m^* and γ^* and then suppressing asterisks), m_0 and m_1 do not represent exactly the ordinary masses of our two bodies (which had been introduced in **2**), but truly these masses, slightly altered by the constant factor

$$1 + (2 + 3\mathfrak{p})\mathfrak{e}.$$

What essentially matters is that they, like their sum m, behave as constants driving the motions to be now considered.

Obviously, the first term in (I) represents the Newtonian attraction, while the other two (both of the second order) are the relativistic perturbations consisting of central attractions; the one varying according the inverse cube, the other according the inverse fourth power of the distance. For the Einsteinian case of one-centre problem, one has only to put $\mathfrak{p} = 0$, and (I) gives the well known expression $3\gamma^2$ for the perturbative function.

4. Orbits described under central forces were thoroughly investigated in the 18th and 19th centuries. Especially for orbits which may be regarded as disturbed Keplerian ellipses, computations of apsidal angles, and corresponding precessions of perihelia may be obtained by elementary methods. In this way, with first order accuracy, the angular precession (per revolution) of the perihelion or, in the case of a double star, of the periastron, is found to be

$$\text{(III)} \qquad \sigma = \sigma_e = 6\pi\mathfrak{a}^2.$$

With the value (7′) of $\mathfrak{a}^2$, namely

$$\frac{fm}{c^2}\,\frac{1}{a(1-e^2)},$$

the expression $\sigma_e = 6\pi\mathfrak{a}^2$ *of* σ *is exactly the precession predicted by Einstein for an infinitesimal planet* ($\mathfrak{p}=0$) *in the case of motion about a central mass possessing the total mass* m *of the binary system.* Therefore, within the required approximation, Einstein's formula, first established for an infinitesimal body in the (relativistic) field of a central mass m, is still valid for two bodies of any masses m_0, m_1 whose sum is m.

I hope that, for some double stars, the precession of the periastron of the satellite star may be observed with sufficient accuracy to test the theoretical result, thus affording a new astronomical confirmation of Einstein's gravitational theory. For the moment I can only draw attention to the matter.

5. The above prediction refers to *relative* orbits. Another, and perhaps more striking, theoretical deduction concerns the *absolute* motion in the sky of any double star system.

It is well known that general relativity does not include, as a rigorous law, the principle of reaction nor its most popular dynamical consequence, concerning the motion of the center of mass in case of absence of external forces. Accordingly, we can no longer rely upon the constance of the absolute velocity $\dot{G} = dG/dx^0$ of the centre of mass G of a double star, but are, on the other hand, enabled to infer the expression of its instantaneous (absolute) acceleration by employing the relativistic treatment of the two-body problem.

The general idea is obvious. Having the Lagrangian equations of motion for the two bodies P_0 and P_1, equations obtained in my previous paper, we deduce at once the vectors $\dot{\boldsymbol{\beta}}_0$ and $\dot{\boldsymbol{\beta}}_1$ of the absolute accelerations as functions of their *relative* positions, and (still) absolute velocities.

The combination

$$\alpha = \frac{1}{m}\,(m_0\boldsymbol{\beta}_0 + m_1\boldsymbol{\beta}_1)$$

is precisely the velocity of G when referred to x^0 as the time variable, i. e., the ordinary velocity divided by c; while the acceleration of G, again referred to x^0, is

$$\dot{\alpha} = \frac{1}{m}\,(m_0\dot{\boldsymbol{\beta}}_0 + m_1\dot{\boldsymbol{\beta}}_1)\,, \tag{14}$$

where the dot denotes d/dx^0.

In view of the classical mechanics, which always holds in the first approximation, we may anticipate that the Newtonian terms in the right-hand member of (14) disappear, so that there remains only the relativistic correc-

tion, expressed, as before, in terms of relative positions and absolute velocities. Now, if we consider actual double star motions, it is plainly permitted, within the degree of approximation in which we are interested, to introduce Newtonian values referring to Keplerian motions of negative energy.

In order to perform the computation along the lines indicated above, it will be convenient to use relative coördinates of invariable direction and having their origin in P_0, where m_0 is the *principal star* ($m_0 \geqq m_1$), and to choose the orthogonal trihedron $P_0x^1x^2x^3$ in its standard position: P_0x^1 towards the periastron of the (undisturbed) elliptical orbit of P_1; P_0x^2 in the plane of this orbit and rotated 90° in the sense of the motion; P_0x^3 forming a right-handed trihedron with the preceding two.

First we recognize from the outlined formulae that *the component* $\dot{\alpha}_3 = 0$. *Therefore, the acceleration of the center of mass* G *of a double star lies entirely in the plane of* (relative) *orbit*; we may also say that it lies in the common plane of (absolute) *orbits*, described by P_0 and P_1 about G.

For the two *components*

$$\dot{\alpha}_1 = \frac{d\alpha_1}{dx^0}, \qquad \dot{\alpha}_2 = \frac{d\alpha_2}{dx^0}$$

in the orbital plane I have found

$$\frac{d\alpha_i}{dx^0} = \mathfrak{p}\mathfrak{d}\left\{ -\frac{d}{dx^0}(\gamma\beta_i) + \frac{\partial\gamma}{\partial x^i}\left(\mathfrak{e} + 4\gamma - \tfrac{3}{2}\frac{1}{\mathfrak{a}^2}\gamma^2\right)\right\}, \quad (i = 1, 2), \tag{15}$$

where the factor

$$\mathfrak{d} = \frac{m_0 - m_1}{m} \tag{16}$$

is proportional to the difference of the masses, while β_i and γ have the same significance as in (3), $\mathfrak{p}$ being defined by (6), $\mathfrak{a}$ by (7), while $\mathfrak{e}$ is the (negative) total energy of the undisturbed Keplerian motion.

6. As already remarked, Keplerian values when used in the right-hand members of (15) give sufficient accuracy. Then the explicit determination of the two variable components of velocity, $\alpha_i(x^0)$, requires only quadratures, easily performed by introducing the true anomaly θ instead of x^0 by means of the relation

$$r^2\frac{d\theta}{dx^0} = \frac{C}{c} = \frac{\sqrt{fma(1-e^2)}}{c}.$$

and remembering that

$$\gamma = \frac{fm}{c^2}\,\frac{1}{r} = \frac{fm}{c^2}\,\frac{1 + e\cos\theta}{a(1-e^2)}.$$

Periodical terms in θ correspond to small fluctuations in the components α_1 and α_2, fluctuations which are repeated during every revolution and certainly remain within the limits of accuracy of observation in the case of all the known double stars. Accordingly, the only interesting terms are the *secular terms,* whose effects accumulate during the successive revolutions. Now we must remember that α_i mean components of velocity with respect to the (Roemerian) time $x^0 = ct$. Therefore *the components of the ordinary velocity of G are $c\alpha_i$.*

Denoting by $c\bar{\alpha}_1$ and $c\bar{\alpha}_2$ their secular parts in terms of θ, we obtain finally, in virtue of (7'),

$$c\bar{\alpha}_1 = -\tfrac{1}{2}\mathfrak{p}\mathfrak{d}\frac{e}{(1-e^2)^{3/2}}\,\frac{fm}{c^2}\sqrt{\frac{fm}{a^3}}\,\theta, \qquad c\bar{\alpha}_2 = 0.$$

The final conclusion is that *the secular acceleration of the center of mass G of the double star is directed along the major axis towards the periastron of the principal star.* The amount of this secular acceleration may be conveniently expressed as the increase of velocity in Km/sec per revolution.

To this end, we first introduce the mass of the Sun, $m_\odot$, and write

$$\frac{fm}{c^2} = \frac{m}{m_\odot}\,\frac{fm_\odot}{c^2}.$$

The second factor is a length, the so called *gravitational radius* $l_\odot$ of the Sun, having a value of the order of magnitude of a Kilometer, or about 1, 5 Km. We may then write

$$\frac{fm}{c^2} = \frac{m}{m_\odot}\cdot 1,5\ \text{Km}.$$

On the other hand, the mean motion $\sqrt{\dfrac{fm}{a^3}}$ of the double star is $\dfrac{2\pi}{T}$, where T is the period of revolution. Of course, T refers to the unit of time used previously in f and c. Starting with the C. G. S. units, we have T in seconds. But, in data of double stars, T is generally expressed in days (for spectroscopic binaries) or in years (for visual binaries). Choosing the first case and writing T^d to avoid ambiguity, we have

$$\frac{1}{T} = \frac{1}{T^d}\,\frac{1}{86164}.$$

It follows then, putting $\theta = 2\pi$ in the preceding expression of $c\bar{\alpha}_1$ and considering its absolute value V, that the increase ΔV of the velocity of G during a revolution is

$$\text{(IV)} \qquad (\Delta V)_{\text{per revolution}} = \tfrac{1}{2}\mathfrak{p}\mathfrak{d}\,\frac{e}{(1-e^2)^{3/2}}\,\frac{m}{m_\odot}\,\frac{4\pi^2}{86164}\,\frac{1,5}{T^d}\,\text{Km/sec.}$$

The number of revolutions per day is $1/T^d$, and, per century, $100 \cdot 365,25$ times $1/T^d$. Since

$$\tfrac{1}{2}\cdot\frac{4\pi^2}{86164}\cdot 1,5\cdot 100\cdot 365,25 = 12,55,$$

the increase of the velocity of the center of mass during a century is

$$\text{(V)} \qquad (\Delta V)_{\text{in a century}} = 12,55\mathfrak{p}\mathfrak{d}\,\frac{e}{(1-e^2)^{3/2}}\,\frac{m}{m_\odot}\,\frac{1}{(T^d)^2}\,\text{Km/sec.}$$

7. Such a difference of velocity along the apsidal line, having a component also in the line of sight, ought to be detectable eventually by spectroscopic observation.

As far as numerical values are concerned, we recognize from (V) [or from (IV)] that the most favorable circumstances are realized for double stars having the following properties:

a) short period, i. e., stars very near each other, which strongly influences $1/T^d$;
b) total mass m, large (or at least not too small) in comparison with the mass of the Sun $m_\odot$;
c) pronounced eccentricity, on account of the factor $e/(1-e^2)^{3/2}$;
d) comparable masses, but not nearly equal, owing to the factors

$$\mathfrak{p} = \frac{m_0 m_1}{(m_0+m_1)^2}, \qquad \mathfrak{d} = \frac{m_0-m_1}{m_0+m_1}.$$

The best conditions for $\mathfrak{p}\mathfrak{d}$ are realized by mass-ratios

$$\frac{m_1}{m_0+m_1} = x, \qquad \frac{m_0}{m_0+m_1} = 1-x$$

for which the polynomial

$$\mathfrak{p}\mathfrak{d} = x(1-x)(1-2x)$$

attains its maximum. This takes places for $x = \tfrac{1}{2}(1-3^{-\frac{1}{2}})$, or, roughly, for two stars containing respectively ¼ and ¾ of the total mass of the system. The corresponding value of $\mathfrak{p}\mathfrak{d}$ is about 0,1.

I am well aware of the astronomical observations which have shown that, in general, the eccentricity e decreases with T^d; so it will not be easy to find a

binary for which the two requirements a) and c) are equally well satisfied. On the other hand, a) predominates in (V), $1/T^d$ appearing to the second power; furthermore, b) is, in the main, in accordance with a).

As the visual binaries have, in general, long periods (some years), they are not to be expected, on account of a), to be advantageous for testing the formula (V). This formula requires, however, the knowledge of the masses of the principal and the companion star. Accordingly, it will be advisable to turn to the class of binaries for which photometric as well as spectrographic observations are available. In order to consider, at least, one example with a reasonable $(\Delta V)_{\text{in a century}}$, I have looked into Moore's Tables of the Lick Observatory,[4] stopping at N° 28, b^1 Persei, for which, unfortunately, only spectroscopic data are certain.

The tabulated elements are

$$m_0 = \frac{0,85}{\sin^3 i} m_\odot, \qquad m_1 = \frac{0,23}{\sin^3 i} m_\odot, \qquad T^d = 1,52, \qquad e = 0,22,$$

i being the *hitherto unknown* inclination of the orbital plane to the tangential plane of the celestial sphere. Whatever i may be, we have finally for b^1 Persei

$$\frac{m}{m_\odot} = \frac{1,08}{\sin^3 i},$$

while $\mathfrak{p}$ and $\mathfrak{d}$, involving only mass-ratio, are independent of i. Their product has the value

$$\mathfrak{p}\mathfrak{d} = 0,09622.$$

Accordingly, formula (V) gives for b^1 Persei

$$(\Delta V)_{\text{in a century}} = 12,55 \cdot 0,09622 \cdot \frac{0,22}{(0,9516)^{3/2}} \, \frac{1,08}{\sin^3 i} \, \frac{1}{(1,52)^2} \text{ Km/sec}$$
$$= \frac{0,13}{\sin^3 i} \text{ Km/sec.}$$

The presence of the (yet unknown) divisor $\sin^3 i$ is consistent with the hope that ΔV may become appreciable much earlier than in a century; perhaps even in a few years.

UNIVERSITY OF ROME.

[4] Or rather to the Resumé, reported in Armellini's *Astronomia siderale*, vol. II (Bologna, Zanichelli, 1931), Appendix 3.

FINITE DEFORMATIONS OF AN ELASTIC SOLID.*

By F. D. Murnaghan.

Introduction. In the classical theory of elasticity a deformation (= strain) is termed infinitesimal when the space derivatives of the components of the displacement vector of an arbitrary particle of the medium are so small that their squares and products may be neglected. Many attempts have been made to extend the classical theory of infinitesimal strain to the case of finite strains i. e. strains in which the fundamental hypothesis which serves to define an infinitesimal strain is not legitimate. The more important of these are given in the references numbered **1** to **7** at the end of the present paper. A good summary with many references may be found in the address of Professor Signorini (**8**) at the Palermo meeting (1935) of the *Societa Italiana per il progresso delle scienze.* In the case of a finite strain there are two essentially different viewpoints which coalesce when the strain is infinitesimal: we may use as the independent variables in terms of which the strain is described either

(a) the coördinates of a typical particle of the medium in the initial or unstrained position or

(b) the coördinates of a typical particle in the final or strained position. Adopting the terminology familiar in the corresponding situation in hydrodynamics we refer to these as the Lagrangian and Eulerian viewpoints respectively. Most of the previous writers on the subject of finite strain have, probably for reasons of mathematical convenience, adopted the Lagrangian view-point but in the present paper (which is concerned with actual applications of the theory) the Eulerian point of view is regarded as fundamentally more significant than the Lagrangian. In this connection the following quotation from the recent paper by Seth (**7**) is to the point:

> "Like the body-stress equations these (the strain components) should be referred to the actual position of a point P of the material in the strained condition, and not to the position of a point considered before strain. The importance of this point, overlooked by various authors, can not be exaggerated. Apparently Filon and Coker (**9**) were the first to notice it and to stress its importance."

In the classical theory one of the fundamental results (derived from the principle of energy conservation) expresses the connection between stress and strain as follows:

* Received February 23, 1937.

The stress tensor equals the gradient of the elastic-energy-density with respect to the strain tensor.

This fundamental principle (which is the formulation of Hooke's law in its most general form) merely states that, in a virtual displacement of the strained elastic medium, the virtual work of all the forces, both surface and body, acting upon the medium may be obtained by integrating over the medium the scalar product of the stress tensor by the variation of the strain tensor. We show in the present paper that this principle is merely an approximation which, whilst valid in the infinitesimal theory, is not valid in the finite theory. The exact principle is that the virtual work is obtained by integrating over the medium the scalar product of the stress tensor by the *space-derivative of the virtual displacement vector* and it is only in the infinitesimal theory that one may with propriety equate the variation of the strain tensor to the space derivative of the virtual displacement vector. It is a fortunate circumstance, the demonstration of which is the raison d'être of the present paper, that the exact equations, valid for any deformation, are sufficiently simple, at least in the case of an isotropic solid, to be applied and to be compared with experimental results. We apply them to Bridgman's experiments with solids and liquids under high pressures (up to 20,000 atmospheres) and find remarkable agreement without introducing more than the two elastic constants of the infinitesimal theory. We also treat the Young's modulus experiment and obtain at least a qualitative explanation of the yield point phenomenon which is not cared for in the classical theory. The mathematical treatment proceeds most naturally and simply when one uses the methods of tensor analysis. It is, however, not necessary to be especially familiar with these methods in order to understand the reasoning and we shall indicate at appropriate places how a non-tensor argument (called, for brevity, Cartesian) would proceed. In the interest of clarity (and probably also of brevity) it has seemed better to make the paper self-contained.

1. The strain tensor and its variation. We are concerned with a three-dimensional medium (which we shall regard as a collection of particles) and with two positions of this medium to which we shall refer as the initial or unstrained and the final or strained position. A typical particle of the medium will have initial and final coördinates. In the classical theory it has been usual to employ the same reference frame (rectangular Cartesian) for both the initial (unstrained) and final (strained) positions and to denote the initial coördinates by (a, b, c) and the final coördinates (of the *same* particle) by (x, y, z). For the treatment we propose here it is inconvenient to tie our hands

at the very beginning by the restrictive hypothesis that the same coördinate reference frame will be used to describe both positions of the medium; the advantage of not making this hypothesis being that it is then possible to make, or contemplate making, transformations of the final coördinates without thereby enforcing a change of the initial coördinates. In the language of tensor analysis the initial coördinates will be invariants or scalars under transformations of the final coördinates. In order to emphasize this fact in the symbolism we shall methodically write the labels necessary to distinguish from one another the various members of a set of scalar quantities to the *left* of the letter which is the symbol for the set; reserving, as is usual, the *right* of a letter for the labels which are necessary to distinguish from one another the various components of a tensor of which the letter is the symbol. Thus we shall denote the initial coördinates of a typical particle of the medium by ${}^{r}a$ and the final coördinates of the same particle by x^s (the labels r and s running independently over the range 1, 2, 3). The initial coördinates ${}^{r}a$, as well as the final coördinates x^s, are, independently of each other, any sets of coördinates which we may find convenient; e. g. the initial coördinates may be rectangular Cartesian and the final coördinates space polar. We make the usual assumption that either coördinate system is differentiable (with continuous first derivatives) with respect to the other and we adopt the notations

$$ {}^{r}a_{,s} = \partial\, {}^{r}a/\partial x^s ; \qquad {}_{s,}x^r = \partial x^r/\partial\, {}^{s}a $$

for the first order partial derivatives. As the notation implies ${}^{r}a_{,s}$ is, for fixed r and varying s, a covariant vector, namely the gradient of the scalar function ${}^{r}a$ whilst ${}_{s,}x^r$ is, for fixed s and varying r, a contravariant vector (which furnishes the direction of the coördinate line along which ${}^{s}a$ varies, the other two a's being held constant). The fundamental reciprocal nature of these two vectors is described by the formulae

$$ ({}^{r}a_{,\sigma})\,({}_{s,}x^{\sigma}) = {}^{r}{}_{s}\delta ; \qquad ({}_{\sigma,}x^{r})\,({}^{\sigma}a_{,s}) = \delta_s{}^{r}. $$

In these formulae we follow the standard convention of tensor analysis according to which a repeated label (in this paper always taken from the Greek alphabet) occurring once above and once below indicates summation with respect to that label over the range 1, 2, 3; and ${}^{r}{}_{s}\delta$, $\delta_s{}^{r}$ each have the value unity when $r = s$ and the value zero otherwise. As the notation implies ${}^{r}{}_{s}\delta$ is a set of 9 scalar functions whilst $\delta_s{}^{r}$ are the 9 components of a single mixed tensor.

The initial and final squared elements of arc length will be given by formulae of the type:

$$ (ds_0)^2 = {}_{\alpha\beta}c\,(d\,{}^{\alpha}a)\,(d\,{}^{\beta}a) ; \qquad (ds)^2 = g_{\alpha\beta}\, dx^{\alpha}\, dx^{\beta} $$

which are induced by the postulated underlying Euclidean metric of the space in which our medium is being deformed. For example if both coördinate systems are rectangular Cartesian

$$(ds_0)^2 = (da)^2 + (db)^2 + (dc)^2; \qquad (ds)^2 = (dx)^2 + (dy)^2 + (dz)^2$$

whilst if the initial coördinate system is rectangular Cartesian and the final space polar

$$(ds_0)^2 = (da)^2 + (db)^2 + (dc)^2; \quad (ds)^2 = (dr)^2 + r^2(d\theta)^2 + r^2 \sin^2\theta (d\phi)^2.$$

On replacing $d\,{}^k a$ by its equivalent: $d({}^k a) = {}^k a_{,\sigma}\, dx^\sigma$ we may express $(ds_0)^2$ in terms of the differentials of the final coördinates x^r:

$$(ds_0)^2 = {}_{\alpha\beta}c\,(d\,{}^\alpha a)\,(d\,{}^\beta a) = h_{\sigma\tau}\, dx^\sigma\, dx^\tau$$

where

$$h_{pq} = {}_{\alpha\beta}c\,({}^\alpha a_{,p})\,({}^\beta a_{,q}).$$

In a similar manner we may express $(ds)^2$ in terms of the differentials of the initial coördinates ${}^r a$:

$$(ds)^2 = g_{\alpha\beta}\, dx^\alpha\, dx^\beta = {}_{\alpha\beta}k\,(d\,{}^\alpha a)\,(d\,{}^\beta a)$$

where

$${}_{pq}k = g_{\alpha\beta}\,({}_{p,}x^\alpha)\,({}_{q,}x^\beta).$$

We obtain thus two equivalent expressions for the difference $(ds)^2 - (ds_0)^2$ of the initial and final squared elements of arc length:

$$(ds)^2 - (ds_0)^2 = 2\epsilon_{\alpha\beta}\, dx^\alpha\, dx^\beta = 2\ {}_{\alpha\beta}\eta\,(d\,{}^\alpha a)\,(d\,{}^\beta a)$$

where

$$2\epsilon_{pq} = g_{pq} - h_{pq}; \qquad 2\ {}_{pq}\eta = {}_{pq}k - {}_{pq}c$$

For a displacement in which lengths are preserved the difference of the squared elements of arc length is zero (identically in the differentials $d\,{}^r a$, or, equivalently, the differentials dx^r) and so the quantities ϵ_{pq} and ${}_{pq}\eta$ are zero for such a *rigid* displacement. In general we regard the quantities ϵ_{pq} or ${}_{pq}\eta$ as descriptive of the strain or deformation and we thus have two methods of describing the strain:

(a) the description, by means of the quantities ϵ_{pq}, in which the final coördinates x^r are adopted as the independent variables in terms of which the description is made and

(b) the description, by means of the quantities ${}_{pq}\eta$, in which the initial coördinates ${}^r a$ are adopted as the independent variables in terms of which the description is made.

In the terminology of the introduction these are, respectively, the Eulerian and Lagrangian descriptions of the strain. We shall also refer to them as the tensor and scalar descriptions, respectively; and shall call the quantities ϵ_{pq} the tensor strain-components and the quantities ${}_{pq}\eta$ the scalar strain-components. The quantities ϵ_{pq} are the covariant components of the strain tensor; this tensor may also be presented in mixed form, or in contravariant form, by means of the formulae

$$\epsilon_s{}^r = g^{ra}\,\epsilon_{as}\,; \qquad \epsilon^{rs} = g^{ra}\,\epsilon_a{}^s$$

where g^{rs} is the reciprocal of g_{rs} : $g^{ra}\,g_{as} = \delta_s{}^r$. Similarly it is convenient to introduce other (but equivalent) scalar descriptions of the strain by means of the formulae:

$${}^p{}_q\eta = ({}^{pa}c)\,({}_{aq}\eta)\,; \qquad {}^{pq}\eta = ({}^{pa}c)\,({}^q{}_a\eta)$$

where ${}^{pq}c$ is the reciprocal of ${}_{pq}c$: $({}^{pa}c)\,({}_{aq}c) = {}^p{}_q\delta$. In the technical language of tensor analysis we may say that we use the tensor g_{pq}, and its reciprocal g^{pq}, for stepping labels down and up upon tensor quantities; and the matrix ${}_{pq}c$, and its reciprocal ${}^{pq}c$, for stepping labels down and up upon scalar sets.

When rectangular Cartesian coördinates, relative to the same reference frame, are used for both the initial and final positions the tensor strain components take the form

$$2\epsilon_{xx} = 1 - \left(\frac{\partial a}{\partial x}\right)^2 - \left(\frac{\partial b}{\partial x}\right)^2 - \left(\frac{\partial c}{\partial x}\right)^2;$$
$$2\epsilon_{yz} = -\frac{\partial a}{\partial y}\frac{\partial a}{\partial z} - \frac{\partial b}{\partial y}\frac{\partial b}{\partial z} - \frac{\partial c}{\partial y}\frac{\partial c}{\partial z} \quad \text{etc.,}$$

whilst the scalar strain components are given by

$$2{}_{aa}\eta = \left(\frac{\partial x}{\partial a}\right)^2 + \left(\frac{\partial y}{\partial a}\right)^2 + \left(\frac{\partial z}{\partial a}\right)^2 - 1\,;$$
$$2{}_{bc}\eta = \frac{\partial x}{\partial b}\frac{\partial x}{\partial c} + \frac{\partial y}{\partial b}\frac{\partial y}{\partial c} + \frac{\partial z}{\partial b}\frac{\partial z}{\partial c} \quad \text{etc.}$$

On denoting the displacement vector $(x-a, y-b, z-c)$ by (u, v, w) we have

$$\epsilon_{xx} = \frac{\partial u}{\partial x} - \tfrac{1}{2}\left\{\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial w}{\partial x}\right)^2\right\};$$
$$\epsilon_{yz} = \tfrac{1}{2}\left(\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}\right) - \tfrac{1}{2}\left\{\frac{\partial u}{\partial y}\frac{\partial u}{\partial z} + \frac{\partial v}{\partial y}\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}\frac{\partial w}{\partial z}\right\} \quad \text{etc.,}$$
$${}_{aa}\eta = \frac{\partial u}{\partial a} + \tfrac{1}{2}\left\{\left(\frac{\partial u}{\partial a}\right)^2 + \left(\frac{\partial v}{\partial a}\right)^2 + \left(\frac{\partial w}{\partial a}\right)^2\right\};$$
$${}_{bc}\eta = \tfrac{1}{2}\left(\frac{\partial v}{\partial c} + \frac{\partial w}{\partial b}\right) + \tfrac{1}{2}\left\{\frac{\partial u}{\partial b}\frac{\partial u}{\partial c} + \frac{\partial v}{\partial b}\frac{\partial v}{\partial c} + \frac{\partial w}{\partial b}\frac{\partial w}{\partial c}\right\} \quad \text{etc.}$$

In the classical infinitesimal theory the partial derivatives $\partial u/\partial x, \cdots, \partial u/\partial a, \cdots$ are regarded as infinitesimal and so we may put

$$\epsilon_{xx} = \frac{\partial u}{\partial x}; \qquad \epsilon_{yz} = \tfrac{1}{2}\left(\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}\right) \text{ etc.,}$$
$$_{aa}\eta = \frac{\partial u}{\partial a}; \qquad _{bc}\eta = \tfrac{1}{2}\left(\frac{\partial v}{\partial c} + \frac{\partial w}{\partial b}\right) \text{ etc.}$$

Since

$$\frac{\partial u}{\partial a} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial a} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial a} + \frac{\partial u}{\partial z}\frac{\partial z}{\partial a} = \frac{\partial u}{\partial x}\left(1 + \frac{\partial u}{\partial a}\right) + \frac{\partial u}{\partial y}\frac{\partial v}{\partial a} + \frac{\partial u}{\partial z}\frac{\partial w}{\partial a}$$

we have, to the degree of approximation contemplated by the infinitesimal theory, $\partial u/\partial a = \partial u/\partial x$ etc., and there is no distinction between the tensor strain components and the scalar strain components.

Relation between the elements of volume. A region occupied by the medium in its initial position is described by setting the coördinates $^r a$ functions of some convenient three independent variables and then the initial element of volume dV_0 is given by the formula $dV_0 = \sqrt{c} \mid d(a) \mid$ where c denotes the determinant of the matrix $_{pq}c$ and $\mid d(a) \mid$ denotes the numerical value of the product of the differentials of the independent variables by the Jacobian determinant of the three coördinates $^r a$ relative to the independent variables. Similarly dV, the element of volume occupied by the *same* particles when in the strained position, is given by the formula $dV = \sqrt{g} \mid d(x) \mid$ where g denotes the determinant of the tensor g_{rs}. Now the relation

$$_{\alpha\beta}c\,(d^\alpha a)\,(d^\beta a) = (ds_0)^2 = h_{\alpha\beta}\, dx^\alpha\, dx^\beta$$

implies

$$\sqrt{c} \mid d(a) \mid = \sqrt{h} \mid d(x) \mid$$

where h denotes the determinant of the tensor h_{rs}. On writing this tensor in its mixed form: $h_{rs} = g_{r\alpha}\, h_s{}^\alpha$ and using the theorem that the determinant of the product of two matrices equals the product of the determinants of the two factors we find $h = g \det(h_s{}^r)$ and so

$$\begin{aligned} dV_0/dV &= \sqrt{c} \mid d(a) \mid / \sqrt{g} \mid d(x) \mid \\ &= \sqrt{h} \mid d(x) \mid / \sqrt{g} \mid d(x) \mid = \sqrt{h/g} = \sqrt{\det(h_s{}^r)}. \end{aligned}$$

On writing the relation $h_{pq} = g_{pq} - 2\epsilon_{pq}$ in its mixed form $h_p{}^q = \delta_q{}^p - 2\epsilon_q{}^p$ we find $\det(h_s{}^r) = 1 - 2I_1 + 4I_2 - 8I_3$ where

$$I_1 = \epsilon_\alpha{}^\alpha; \qquad I_2 = \frac{1}{2!}\,\delta^{\alpha_1\alpha_2}_{\beta_1\beta_2}\epsilon_{\alpha_1}{}^{\beta_1}\epsilon_{\alpha_2}{}^{\beta_2}; \qquad I_3 = \frac{1}{3!}\,\delta^{\alpha_1\alpha_2\alpha_3}_{\beta_1\beta_2\beta_3}\epsilon_{\alpha_1}{}^{\beta_1}\epsilon_{\alpha_2}{}^{\beta_2}\epsilon_{\alpha_3}{}^{\beta_3}$$

are the three strain invariants. They are, respectively, the sum of the diagonal elements, the sum of the principal two rowed minors, and the determinant of the matrix $\epsilon_q{}^p$ which presents the strain tensor in mixed form. When the strain is homogeneous i. e. does not vary from point to point of the medium we may replace the ratio dV_0/dV by the ratio V_0/V where V_0 is the initial volume of any portion of the medium and V is the volume occupied in the strained position, by the particles which initially occupied the volume V_0. We have, then, for a homogeneous strain the relation

$$V_0/V = \sqrt{\det.(\delta_s{}^r - 2\epsilon_s{}^r)} = \sqrt{1 - 2I_1 + 4I_2 - 8I_3}.$$

For the special case of a homogeneous strain which is also at each point isotropic (which will be the case in an isotropic medium subjected to uniform hydrostatic pressure) the strain tensor will be a scalar tensor: $\epsilon_s{}^r = \epsilon\delta_s{}^r$ and we have the relatively simple relation

$$V_0/V = (1 - 2\epsilon)^{3/2}.$$

For the classical infinitesimal theory this reduces, as is at once seen on writing $(1 - 2\epsilon)^{3/2} = 1 - 3\epsilon + \cdots$, to $V_0/V = 1 - 3\epsilon$ or

$$\epsilon = \tfrac{1}{3} \cdot \frac{V - V_0}{V} = \tfrac{1}{3}\frac{\Delta V}{V} = \tfrac{1}{3}\frac{\Delta V}{V_0}$$

(to the degree of approximation contemplated by the theory). The exact relation, valid for a finite strain, to which the relation just written is an approximation is

$$\epsilon = \tfrac{1}{2}\{1 - (V_0/V)^{2/3}\}.$$

The mathematical description of a homogeneous strain is that the strain tensor should be constant *in the tensor sense* i. e. its absolute or covariant derivative vanishes. This implies that the absolute derivatives of the invariants I_1, I_2, I_3 (or what is the same thing, since these are scalars, their space derivatives) vanish so that dV_0/dV is a numerical constant. The hypothesis $\epsilon_{pq,r} = 0$ (where as is usual labels following a comma indicate covariant differentiation) implies $h_{pq,r} = 0$ and since $h_{pq} = ({}_{\alpha\beta}c)\, {}^{\alpha}a_{,p}\, {}^{\beta}a_{,q}$ an easy calculation yields ${}^{k}a_{,pq} + ({}^{k}{}_{\alpha\beta}\Gamma)\, {}^{\alpha}a_{,p}\, {}^{\beta}a_{,q} = 0$. In particular when the a's and x's are rectangular Cartesian coördinates so that ${}^{k}a_{,pq} = \partial^2\, {}^{k}a/\partial x^p\, \partial x^q$ and ${}^{k}{}_{rs}\Gamma = 0$ the ${}^{k}a$ must be linear functions of the x^r. Conversely if this is the case the strain components are numerical constants and the strain is homogeneous.

The variation of the strain tensor. In order to introduce the concept of virtual work and thereby express the conditions for equilibrium of the strained medium we must adopt the dynamic as opposed to the static viewpoint. In other words instead of regarding the strained position of the medium as something fixed and final we must regard it as capable of variation. To do this

conveniently, from the mathematical viewpoint, we conceive of the final coördinates x^s as depending not only upon the initial coördinates ${}^r a$ but also upon an accessory parameter θ (which could, in hydrodynamics, conveniently be taken as the time variable). We shall denote differentials with respect to the parameter θ by the symbol D:

$$Dx^s = \frac{\partial x^s}{\partial \theta} d\theta$$

it being understood that in the partial differentiation with respect to θ the coördinates ${}^r a$ are kept constant (i. e. the D denotes the substantial or particle differentiation of hydrodynamics). If we have any tensor function $f^{\cdots}_{\cdots\cdot}$ of the coördinates x^r we shall denote by $\delta f^{\cdots}_{\cdots\cdot}$ the tensor of the same type defined by the rule

$$\delta f^{\cdots}_{\cdots\cdot} = f^{\cdots}_{\cdots\cdot, a} Dx^a.$$

When the coördinates x^r are Cartesian covariant differentiation is merely ordinary space differentiation of the tensor components and so $\delta f^{\cdots}_{\cdots\cdot} = Df^{\cdots}_{\cdots\cdot}$. Even when the coördinates are not Cartesian this relation holds for any scalar function: $\delta f = Df$ since $f_{,r} = \partial f / \partial x^r$. We effect a small economy of notation by defining δx^r by the relation $\delta x^r = Dx^r$ and we refer to the contravariant vector δx^r as the virtual displacement vector. Before proceeding to the necessary calculation of the variation of the strain tensor and of its scalar components it must be clearly understood that the variation of any scalar function of the ${}^r a$ is zero; e. g. $\delta_{pq} c = 0$. This result is a trivial consequence of the definition of the δ symbol for in the differentiation with respect to θ the coördinates ${}^r a$ are held constant.

Owing to the independence of the variables ${}^r a$ and θ differentiations with respect to them are interchangeable as to order (it being supposed that the second order derivatives involved exist and are continuous). Hence

$$D({}_s x^k) = \frac{\partial}{\partial^s a} \delta x^k = \frac{\partial}{\partial x^a} (\delta x^k) \cdot ({}_s x^a)$$

and on multiplication by $d^s a$ (which is independent of θ) and summation with respect to s we find

$$D(dx^k) = \frac{\partial}{\partial x^a} (\delta x^k) \cdot dx^a.$$

This implies the tensor equation

$$\delta(dx^k) = (\delta x^k)_{,a}\, dx^a$$

since it is the form to which the tensor equation reduces when the coördinates

x^k are Cartesian. Since the variation of the metrical tensor is zero: $\delta g_{rs} = 0$, the tensor equation just written may be put in the equivalent form

$$\delta(dx_k) = (\delta x_k)_{,a}\, dx^a.$$

Since $d({}^r a) = {}^r a_{,a}\, dx^a$ and $\delta d({}^r a) = 0$ we have

$$\delta({}^r a_{,a})\, dx^a = -\, {}^r a_{,a}\, \delta(dx^a) = -\, {}^r a_{,a} (\delta x^a)_{,\beta}\, dx^\beta$$

and since this must hold for arbitrary dx^r we must have

$$\delta({}^r a_{,k}) = -\, {}^r a_{,a} (\delta x^a)_{,k}.$$

From this expression and the relation $h_{pq} = {}_{a\beta}c\, ({}^a a_{,p})\, ({}^\beta a_{,q})$ we can read off at once the formula for the variation of h_{pq} (or, what is the same thing, of $-\,2\epsilon_{pq}$). We find

$$\begin{aligned} \delta h_{pq} &= {}_{a\beta}c\, (\delta^a a_{,p} \cdot {}^\beta a_q + {}^a a_{,p}\, \delta^\beta a_{,q}) \\ &= -\, {}_{a\beta}c \{ {}^a a_{,\tau} (\delta x^\tau)_{,p}\, {}^\beta a_q + ({}^a a_{,p})\, {}^\beta a_{,\tau} (\delta x^\tau)_{,q} \} \\ &= -\, h_{\tau q} (\delta x^\tau)_{,p} - h_{p\tau} (\delta x^\tau)_{,q} \\ &= -\, h_q^{\tau} (\delta x_\tau)_{,p} - h_p^{\tau} (\delta x_\tau)_{,q}. \end{aligned}$$

This will be seen to be the significant formula for our later purpose. On writing $h_{pq} = g_{pq} - 2\epsilon_{pq}$ it may be written in the equivalent form

$$\delta\epsilon_{pq} = \tfrac{1}{2}\{ (\delta x_q)_{,p} + (\delta x_p)_{,q} \} - \{ \epsilon_q^{\tau} (\delta x_\tau)_{,p} + \epsilon_p^{\tau} (\delta x_\tau)_{,q} \}.$$

For the classical infinitesimal theory it is allowable to write

$$\delta\epsilon_{pq} = \tfrac{1}{2}\{ (\delta x_q)_{,p} + (\delta x_p)_{,q} \}$$

and it is the difference between this and the exact expression just furnished that makes it incorrect, as stated in the introduction, to write the stress tensor as the gradient with respect to the strain tensor of the elastic energy density.

Criterion for a rigid virtual displacement. A virtual displacement is said to be rigid when $\delta(ds^2) = 0$. Since $\delta(ds_0)^2 = 0$ and

$$(ds)^2 - (ds_0)^2 = 2\epsilon_{a\beta}\, dx^a\, dx^\beta$$

an equivalent description of a rigid virtual displacement is

$$\delta(\epsilon_{a\beta}\, dx^a\, dx^\beta) = 0.$$

One must be on one's guard against the error of supposing that (because in a rigid displacement $\epsilon_{pq} = 0$) in a virtual rigid displacement the strain tensor

ϵ_{pq} is constant: $\delta\epsilon_{pq} = 0$. The evident fallacy in such a guess being the neglect of terms involving $\delta(dx^r)$ which quantities are not in general zero in a rigid displacement. On the other hand the scalar components ${}_{pq}\eta$ of the strain tensor are constant in a rigid virtual displacement; for the criterion for a rigid virtual displacement may be put in the form $\delta({}_{\alpha\beta}\eta\, d\,{}^{\alpha}a\, d\,{}^{\beta}a) = 0$ and this is equivalent to $\delta\,{}_{\alpha\beta}\eta(d\,{}^{\alpha}a)(d\,{}^{\beta}a) = 0$ since $d\,{}^{r}a$ is independent of θ. Since this equation must hold for arbitrary $d\,{}^{r}a$ we must have $\delta_{pq}\eta = 0$ and this necessary condition is clearly sufficient. For any virtual displacement, rigid or not, we have

$$\begin{aligned}\delta(ds)^2 &= \delta(g_{\alpha\beta}\, dx^\alpha\, dx^\beta) = \delta(dx^\alpha\, dx_\alpha)\\ &= dx^\alpha(\delta x_\alpha)_{,\beta}\, dx^\beta + (\delta x^\alpha)_{,\beta}\, dx^\beta\, dx_\alpha\\ &= \{(\delta x_\alpha)_{,\beta} + (\delta x_\beta)_{,\alpha}\} dx^\alpha\, dx^\beta\end{aligned}$$

so that the criterion for a rigid virtual displacement may be put in the form $(\delta x_p)_{,q} + (\delta x_q)_{,p} = 0$ (equations of Killing). Amongst the possible rigid virtual displacements are those for which $(\delta x_p)_{,q} = 0$; these are the virtual translations which appear, when the coördinates x_p are Cartesian, in the form $\delta x_p =$ constant. Since $\delta(ds)^2 = 2\delta({}_{\alpha\beta}\eta\, d\,{}^{\alpha}a\, d\,{}^{\beta}a)$ we have, for any virtual displacement whatsoever,

$$2\delta\,{}_{\alpha\beta}\eta \cdot d\,{}^{\alpha}a \cdot d\,{}^{\beta}a = \{(\delta x_\alpha)_{,\beta} + (\delta x_\beta)_{,\alpha}\} dx^\alpha\, dx^\beta$$

implying

$$\delta\,{}_{pq}\eta = \tfrac{1}{2}\{(\delta x_\alpha)_{,\beta} + (\delta x_\beta)_{,\alpha}\}({}_{p,}x^\alpha)({}_{q,}x^\beta).$$

2. The stress-tensor and the virtual work of the applied forces acting on the medium. Let us consider any portion of our elastic medium which is bounded by a closed surface and denote by S_0 and S the initial and final positions, respectively, of this bounding surface. The coördinates ${}^{r}a$ of the initial position of any particle of this bounding surface (and equally the coördinates x^s of the final position of such a particle) are functions of two independent parameters and the surface element of S may be described by means of the covariant vector $dS_r = \sqrt{g}\, d(x^p, x^q)$ (where p, q, r is an even permutation of the natural order 1, 2, 3 and $d(x^p, x^q)$ denotes the product of the differentials of the two independent variables times the Jacobian of the two coördinates x^p and x^q relative to these independent variables). When the coördinates x^r are rectangular Cartesian

$$dS_x = d(y, z); \qquad dS_y = d(z, x); \qquad dS_z = d(x, y).$$

Denoting by dS the magnitude of this typical surface element—so that

$$(dS)^2 = g^{\alpha\beta}\, dS_\alpha\, dS_\beta$$

—the stress tensor T^{rs} is a linear vector function which associates with each surface element dS_r a stress vector F^r by means of the formula $F^r dS = T^{r\alpha} dS_\alpha$. The virtual work of the stresses across the boundary S is accordingly

$$\int_S (F^\beta\, dS)\,\delta x_\beta = \int_S T^{\beta\alpha}\, dS_\alpha\, \delta x_\beta$$

and this is equivalent to the volume integral

$$\int_V (T^{\beta\alpha}\, \delta x_\beta)_{,\alpha}\, dV$$

extended over the volume V bounded by S. If there are mass forces—M^r per unit mass—acting on the medium the virtual work of these is $\int_V \rho M^\beta\, \delta x_\beta\, dV$, where ρ is the mass density, and so the virtual work of all the forces acting on any portion of the medium is

$$\int_V \{(T^{\beta\alpha}{}_{,\alpha} + \rho M^\beta)\ \delta x_\beta + T^{\beta\alpha}(\delta x_\beta)_{,\alpha}\} dV.$$

We now make the physical assumption (criterion of equilibrium) that this virtual work is zero for any *rigid* virtual displacement. Amongst these rigid virtual displacements are the translations which are characterized by the tensor equation $(\delta x_p)_{,q} = 0$ and so we must have

$$\int_V (T^{\beta\alpha}{}_{,\alpha} + \rho M^\beta)\,\delta x_\beta\, dV = 0;$$

since δx_r may be assigned arbitrarily at a given point and since the volume V of integration is arbitrary this forces

$$T^{r\alpha}{}_{,\alpha} + \rho M^r = 0.$$

Consequently the virtual work of all the forces (mass and surface) acting upon any portion of the medium in any virtual displacement whatever is given by the expression

$$\text{Virtual work} = \int_V T^{\beta\alpha}(\delta x_\beta)_{,\alpha}\, dV.$$

Furthermore since this must vanish for any rigid virtual displacement i. e. for any virtual displacement for which $(\delta x_p)_{,q} + (\delta x_q)_{,p} = 0$ the stress tensor must be symmetric: $T^{sr} = T^{rs}$. This relation enables us to write the expression for the virtual work of all the forces acting upon the medium in any virtual displacement whatever in the form:

$$\text{Virtual work} = \tfrac{1}{2}\int_V T^{\alpha\beta}\{(\delta x_\alpha)_{,\beta} + (\delta x_\beta)_{,\alpha}\}dV.$$

In the classical infinitesimal theory this may be written as $\int T^{\alpha\beta}\,\delta\epsilon_{\alpha\beta}\,dV$ but such an approximation is not legitimate in the finite theory.

3. The elastic potential and its connection with the stress tensor. We now consider an element of volume dV of the medium in its strained position and denote by ρ the density of the matter occupying the element of volume dV so that the element of mass is $dm = \rho\, dV$. The principle of conservation of mass is expressed by the formula:

$$\delta(dm) = \delta(\rho\, dV) = 0.$$

In order to apply the fundamental energy-conservation law of thermodynamics we denote by T the temperature of the element of mass dm; by σ the entropy density (per unit mass) so that the entropy of the mass dm is $\sigma\, dm = \rho\sigma\, dV$; and by udm the internal energy of the mass dm. Then the principle of thermodynamics to which we have referred says that $T\delta(\sigma\, dm) = \delta(udm)$—Virtual work of all forces acting on dm. On introducing the free-energy density $\phi = u - T\sigma$ and availing ourselves of the principle of conservation of mass: $\delta\, dm = 0$ we find, on integrating over any portion V of the strained medium,

$$\int_V \delta\phi\, dm = \int_V T^{\alpha\beta}(\delta x_\alpha)_{,\beta}\, dV - \int_V \sigma\, dm \cdot \delta T.$$

On writing $dm = \rho\, dV$ and observing that this relation must hold for arbitrary volumes V we obtain the fundamental formula connecting the elastic potential ϕ with the stress tensor T^{rs}:

$$\rho\,\delta\phi = T^{\alpha\beta}(\delta x_\alpha)_{,\beta} - \rho\sigma\,\delta T.$$

In order to derive from this a connection between the stress tensor T^{rs} and the strain tensor ϵ_{rs} we must make some hypothesis concerning the function ϕ. We shall assume that it depends only on the three gradient vectors ${}^r a_{,s}$ (or, equivalently, on the reciprocal set ${}_{r,}x^s$) it being understood that either set may appear both covariantly and contravariantly. In other words we shall assume that ϕ is a function of the vectors ${}^r a_{,s}$, the metrical tensor g_{rs}, the scalar quantities ${}_{pq}c$ and the temperature T. We shall confine our attention in what follows to isothermal variations so that T is a constant parameter in ϕ to which attention need not be explicitly directed. Then $\delta\phi$ must be zero in any (isothermal) rigid virtual displacement; in other words $\dfrac{\partial\phi}{\partial({}^\alpha a_{,\beta})}\,\delta({}^\alpha a_{,\beta}) = 0.$

provided $(\delta x_p)_{,q} + (\delta x_q)_{,p} = 0$. On inserting the value $-({}^r a_{,a})(\delta x^a)_{,k}$ given in 1 for $\delta({}^r a_{,k})$ we see that the tensor $\frac{\partial \phi}{\partial({}^a a_{,p})}({}^a a^{,q})$ must be symmetric in p and q $({}^r a^{,q} = g^{qa}({}^r a_{,a}))$. Hence the function ϕ of the nine variables ${}^r a_{,s}$ is conditioned by the three linear, homogeneous, first order, partial differential equations:

$$\frac{\partial \phi}{\partial({}^a a_{,p})}\, {}^a a^{,q} = \frac{\partial \phi}{\partial({}^a a_{,q})}\, {}^a a^{,p}.$$

These equations are readily seen to form a complete system (the commutator of any two being the third) and so the general solution of them is a function of $6 = 9 - 3$ independent solutions. Since $\delta\, {}_{pq}\eta = 0$ in any rigid virtual displacement we know that the six functions ${}_{pq}\eta$ satisfy the three partial differential equations just written. Hence ϕ can involve the vectors ${}^r a_{,s}$ only through the scalar functions ${}_{pq}\eta$: $\phi = \phi({}_{pq}\eta, {}_{pq}c, T)$. The quantities g_{rs} cannot occur in the expression for ϕ since ϕ is a scalar and the only scalar functions of the metrical tensor g_{rs} are numerical (e. g. $g_a{}^a = \delta_a{}^a = 3$). Now under a transformation of coördinates ${}^r a$ the quantities ${}_{pq}\eta$, ${}_{pq}c$ (which are scalars under transformation of coördinates x^s) transform as covariant tensors. We say that the medium is *isotropic* when the elastic-energy density ϕ is unaffected by a transformation of the coördinates ${}^r a$; for instance when the coördinates ${}^r a$ are rectangular Cartesian an arbitrary rotation of the reference frame must leave the function ϕ unaffected. In order that this may be the case ϕ must involve ${}_{pq}\eta$, ${}_{pq}c$ only through the "invariants" (under transformation of the coördinates ${}^r a$):

$$J_1 = {}^a{}_a\eta; \quad J_2 = \tfrac{1}{2}\,\delta^{a_1 a_2}_{\beta_1 \beta_2}\,({}^{\beta_1}{}_{a_1}\eta)({}^{\beta_2}{}_{a_2}\eta); \quad J_3 = \frac{1}{3!}\,\delta^{a_1 a_2 a_3}_{\beta_1 \beta_2 \beta_3}\,({}^{\beta_1}{}_{a_1}\eta)({}^{\beta_2}{}_{a_2}\eta)({}^{\beta_3}{}_{a_3}\eta).$$

These quantities J_1, J_2, J_3 are, respectively, the sum of the diagonal elements, the sum of the principal two-rowed minors, and the determinant of the matrix ${}^p{}_q\eta$. Hence, for an isotropic medium: $\phi = \phi(J_1, J_2, J_3, T)$. We prove in the appendix that the quantities J_1, J_2, J_3 are functions of the three invariants I_1, I_2, I_3 of the stress tensor $\epsilon_s{}^r$ and so we may write, for an isotropic medium,

$$\phi = \phi(I_1, I_2, I_3, T)$$

i. e. ϕ is a function of the components of the strain tensor $\epsilon_s{}^r$ and T. Conversely if we make the hypothesis that ϕ is a function of T and of the strain components (tensor) $\epsilon_s{}^r$ *alone* this implies that ϕ is isotropic; for the only scalar functions of $\epsilon_s{}^r$ are functions of its invariants I_1, I_2, I_3. In order to prevent possible misunderstanding of this remark [in view of the fact that in

the classical treatment of elasticity ϕ is taken for crystalline (= non-isotropic) media as a quadratic function of the strain components] we may say that it is tacitly understood in this classical procedure that a special privileged reference frame, determined by the axes of the crystal, has been chosen. The coefficients of the quadratic form are accordingly not scalar quantities but constitute a tensor which depends on the orientation of the crystalline axes.

The fundamental stress-strain relations for an isotropic medium. As we have just seen the elastic energy density ϕ is, for an isotropic medium, a function of the tensor strain components $\epsilon_s{}^r$, involving these through the strain invariants I_1, I_2, I_3. It will be, for the moment, more convenient to regard ϕ as a function of the covariant strain components ϵ_{rs} ($\epsilon_s{}^r = g^{ra}\,\epsilon_{as}$!) and we shall, if necessary, symmetrize its formal expression; i. e. we shall replace each ϵ_{rs}, wherever it occurs in the expression for ϕ, by its equivalent: $\epsilon_{rs} = \frac{1}{2}(\epsilon_{rs} + \epsilon_{sr})$. Denoting by $\partial\phi/\partial\epsilon_{rs}$ the partial derivative of ϕ with respect to ϵ_{rs} all the other ϵ_{rs} (including ϵ_{sr}) being held constant in the differentiation (so that in this formal differentiation no attention is paid to the symmetry relations $\epsilon_{sr} = \epsilon_{rs}$) it follows that $\partial\phi/\partial\epsilon_{rs} = \partial\phi/\partial\epsilon_{sr}$. We have seen in 1 that

$$\delta\epsilon_{pq} = -\tfrac{1}{2}\delta h_{pq} = \tfrac{1}{2}\{h_q{}^\tau(\delta x_\tau)_{,p} + h_p{}^\tau(\delta x_\tau)_{,q}\}$$

where $h_{pq} = g_{pq} - 2\epsilon_{pq}$, and hence

$$\delta\phi = \frac{\partial\phi}{\partial\epsilon_{\alpha\beta}}\,\delta\epsilon_{\alpha\beta} = \tfrac{1}{2}\,\frac{\partial\phi}{\partial\epsilon_{\alpha\beta}}\,\{h_\beta{}^\tau(\delta x_\tau)_{,\alpha} + h_\alpha{}^\tau(\delta x_\tau)_{,\beta}\} = \frac{\partial\phi}{\partial\epsilon_{\alpha\beta}}\,h_\beta{}^\tau(\delta x_\tau)_{,\alpha}$$

since $\partial\phi/\partial\epsilon_{rs} = \partial\phi/\partial\epsilon_{sr}$. Since in an isothermal virtual displacement $\rho\,\delta\phi = T^{\alpha\beta}(\delta x_\alpha)_{,\beta}$ it follows that, for an isotropic medium,

$$\rho\,\frac{\partial\phi}{\partial\epsilon_{\alpha\beta}}\,h_\beta{}^\tau(\delta x_\tau)_{,\alpha} = T^{\alpha\beta}(\delta x_\alpha)_{,\beta}.$$

Since the virtual displacement is arbitrary the components $(\delta x_p)_{,q}$ of the space derivative of the virtual displacement vector may be assigned arbitrary values at any point x^r and so we must have

$$T^{rs} = \rho\,\frac{\partial\phi}{\partial\epsilon_{s\beta}}\,h_\beta{}^r = \rho\left(\frac{\partial\phi}{\partial\epsilon_{sr}} - 2\epsilon_\beta{}^r\,\frac{\partial\phi}{\partial\epsilon_{s\beta}}\right).$$

That these equations imply $T^{sr} = T^{rs}$ follows from the fact that ϕ involves the strain components ϵ_{rs} only through the strain invariants I_1, I_2, I_3. For example $I_1 = \epsilon_\alpha{}^\alpha = g^{\alpha\beta}\,\epsilon_{\alpha\beta}$ so that

$$\epsilon_\beta{}^r \frac{\partial I_1}{\partial \epsilon_{s\beta}} = \epsilon_\beta{}^r g^{s\beta} = \epsilon^{rs} = \epsilon_\beta{}^s \frac{\partial I_1}{\partial \epsilon_{r\beta}}.$$

The formulae just obtained are fundamental but it is frequently more convenient to present the stress tensor in its mixed form $T_s{}^r$. Since

$$\epsilon_s{}^r = g^{ra} \epsilon_{as}; \quad \frac{\partial \phi}{\partial \epsilon_{rs}} = g^{ra} \frac{\partial \phi}{\partial \epsilon_s{}^a}$$

and so our formulae appear as

$$T_s{}^r = \rho \left(\frac{\partial \phi}{\partial \epsilon_r{}^s} - 2\epsilon_\beta{}^r \frac{\partial \phi}{\partial \epsilon_\beta{}^s} \right).$$

The result $dV_0/dV = \sqrt{1 - 2I_1 + 4I_2 - 8I_3}$, obtained in 1, may be written in the equivalent form $\rho = \rho_0 \sqrt{1 - 2I_1 + 4I_2 - 8I_3}$ (since the principle of conservation of mass implies $\rho \, dV = \rho_0 \, dV_0$). In the classical (infinitesimal) theory the strain invariants I_1, I_2, I_3 are infinitesimal quantities of the first, second and third orders of magnitude respectively. We may therefore write, to a first approximation, $\rho = \rho_0$ and to a second approximation $\rho = \rho_0(1 - I_1)$. Keeping only the first approximation our fundamental stress strain relations reduce to

$$T_s{}^r = \rho_0 \frac{\partial \phi}{\partial \epsilon_r{}^s} = \frac{\partial \phi'}{\partial \epsilon_r{}^s}; \qquad \phi' = \rho_0 \phi.$$

These are the basic formulae (expressing Hooke's law) of the classicical theory, ϕ' being the elastic energy per unit initial volume (or, what is the same thing to the degree of approximation contemplated by the classical theory, per unit final volume).

Even in the case of finite strains the strain invariants are relatively small. The two cases to which we shall devote some attention in detail in the present paper are:

(a) uniform hydrostatic pressure; here the stress and strain tensors are scalar tensors: $T_s{}^r = -p\delta_s{}^r$; $\epsilon_s{}^r = -f\delta_s{}^r$ (where p is what is commonly called pressure) and $I_1 = -3f$, $I_2 = 3f^2$, $I_3 = -f^3$. We shall discuss a little later the agreement of the finite theory with the results of experiments by Bridgman (**10**) on the compressibility of sodium under pressures ranging from 2,000 atmospheres to 20,000 atmospheres. In these experiments the ratio $(V_0 - V)/V_0$ varied from .030 at a pressure of 2,000 atm., to .189 at a pressure of 20,000 atm. Since $V_0/V = \sqrt{\det(\delta_s{}^r - 2\epsilon_s{}^r)} = (1 + 2f)^{3/2}$ the corresponding values of f are .010 and .075 respectively; hence I_1 varied between $-.030$ and $-.225$, I_2 varied between .001 and .0169 whilst I_3 varied between $-.000001$ and $-.00042$.

(b) uniform linear stress (Young's modulus experiment). Here the stress tensor is scalar with two components zero whilst the strain tensor is scalar with two components equal:

$$T_x^x = T_y^y = 0; \qquad \epsilon_x^x = \epsilon_y^y = -\sigma\epsilon_z^z$$

where σ, Poisson's ratio, has a value $< .5$. From the formula giving the volume change we have $V_0/V = (1 + 2\sigma\epsilon_z^z)\sqrt{1 - 2\epsilon_z^z}$ so that ϵ_z^z cannot surpass the value .5. Hence $I_1 = (1 - 2\sigma)\epsilon_z^z$ is a small fraction even for very large strains and $I_2 = (\sigma^2 - 2\sigma)(\epsilon_z^z)^2$, $I_3 = \sigma^2(\epsilon_z^z)^3$ are smaller.

We may, therefore, even for large strains hope to secure good agreement with experiment by expanding $\phi(\epsilon_s^r)$ as a power series in the strain components and neglecting terms of orders of magnitude greater than an agreed upon order (say the second, third, etc.). In the classical, infinitesimal, theory the agreed upon order is the second. We shall agree upon the third but we call explicit attention at this point to the fact that our theory gives remarkable agreement with experimental results even if we do not introduce any more constants (i. e. coefficients in the expansion of the elastic energy density ϕ) than those (two in number) introduced in the infinitesimal theory. In fact in the case of the compressibility experiments the two constants combine into a single one so that a *one* constant formula suffices to predict to a high degree of accuracy the connection between pressure and volume over the extensive range from 2,000 to 20,000 atmospheres. It is clear that an additive constant in ϕ is of no significance since ϕ enters our fundamental equations only through its partial derivatives. In order to keep as closely as possible to the notations of the classical (infinitesimal) theory we shall expand $\rho_0\phi$ instead of ϕ:

$$\rho_0\phi = \alpha I_1 + \frac{\lambda + 2\mu}{2} I_1^2 - 2\mu I_2 + lI_1^3 + mI_1I_2 + nI_3.$$

As we shall see in a moment the hypothesis that the stress is zero in the initial state (characterised by $\epsilon_s^r = 0$) forces $\alpha = 0$ and the assumption of the infinitesimal theory is

$$\phi' = \rho_0\phi = \frac{\lambda + 2\mu}{2} I_1^2 - 2\mu I_2$$

(where, of course, the invariants I_1, I_2 of the classical theory are the approximations obtained by neglecting the second order terms in the expressions for the strain components). In the usual presentations of the infinitesimal theory the invariant $I'_2 = \epsilon_\beta^\alpha\epsilon_\alpha^\beta = I_1^2 - 2I_2$ is used instead of our I_2 and the expression for ϕ' appears as

$$\phi' = \frac{\lambda}{2} I_1{}^2 + \mu I'_2.$$

On using the relations

$$\partial I_1/\partial \epsilon_r{}^s = \delta_s{}^r; \qquad \partial I_2/\partial \epsilon_r{}^s = \delta_s{}^r I_1 - \epsilon_s{}^r \qquad \partial I_3/\partial \epsilon_r{}^s = I_3 \zeta_s{}^r$$

where $\zeta_s{}^r$ is the tensor reciprocal to $\epsilon_s{}^r$:

$$\zeta_a{}^r \epsilon_s{}^a = \delta_s{}^r$$

and remembering that $\rho/\rho_0 = \sqrt{1 - 2I_2 + 4I_2 - 8I_3} = 1 - I_1 + \cdots$ we find

$${}^r = (1 - I_1 + \cdots) \left\{ \begin{array}{l} (\alpha + \lambda I_1 + (3l + m) I_1{}^2 + m I_2 - 2n I_3) \delta_s{}^r \\ + [2(\mu - \alpha) - (m + 2\lambda) I_1 - 2(3l + m) I_1{}^2 - 2m I_2] \epsilon_s{}^r \\ + 2(m I_1 - 2\mu) \epsilon_a{}^r \epsilon_s{}^a + n I_3 \zeta_s{}^r \end{array} \right\}.$$

The value of $T_s{}^r$ in the initial position $(\epsilon_s{}^r = 0)$ is, accordingly

$$(T_s{}^r)_0 = \alpha \delta_s{}^r$$

(so that the assumption of isotropy forces the stress in the initial position to be scalar i. e. of the nature of a hydrostatic pressure); we shall make the usual assumption that the stress in the initial position is zero forcing $\alpha = 0$. On multiplying out by the factor $1 - I_1 + \cdots$ and neglecting quantities of higher order than the second we find

$$\begin{aligned} T_s{}^r = {} & \{I_1 + (3l + m - \lambda) I_1{}^2 + m I_2\} \delta_s{}^r \\ & + \{2\mu - (m + 2\lambda + 2\mu) I_1\} \epsilon_s{}^r - 4\mu \epsilon_a{}^r \epsilon_s{}^a + n I_3 \zeta_s{}^r. \end{aligned}$$

We shall not keep in the following the quantities arising from the third order terms in the expansion of $\rho_0 \phi$ i. e. we shall set $l = 0$, $m = 0$, $n = 0$ when we obtain

$$T_s{}^r = \lambda I_1 (1 - I_1) \delta_s{}^r + 2\{\mu - (\lambda + \mu) I_1\} \epsilon_s{}^r - 4\mu \epsilon_a{}^r \epsilon_s{}^a.$$

The first invariant $T = T_a{}^a$ of the stress tensor is, accordingly,

$$\begin{aligned} T &= 3\lambda I_1 (1 - I_1) + 2\{\mu - (\lambda + \mu) I_1\} I_1 - 4\mu I'_2 \\ &= (3\lambda + 2\mu) I_1 - (5\lambda + 6\mu) I_1{}^2 + 8\mu I_2. \end{aligned}$$

If we neglect the second order terms we obtain the expressions of the infinitesimal theory

$$T_s{}^r = \lambda I_1 \delta_s{}^r + 2\mu \epsilon_r{}^s; \qquad T = (3\lambda + 2\mu) I_1.$$

Although, as we shall see, a good approximation to the results of experiments may be secured by using only two elastic constants λ, μ i. e. by neglecting the

third order terms in the expansion of $\rho_0\phi$ a true second order approximation would keep these third order terms thus introducing five elastic constants λ, μ, l, m, n. Neglecting, however, third order terms in the expressions for the stress tensor the second order approximation is

$$T_s{}^r = \{\lambda I_1 + (3l + m - \lambda) I_1{}^2 + m I_2\}\delta_s{}^r$$
$$+ \{2\mu - (m + 2\lambda + 2\mu) I_1\}\epsilon_s{}^r - 4\mu\epsilon_a{}^r\epsilon_s{}^a + n I_3 \zeta_s{}^r$$
$$T = (3\lambda + 2\mu) I_1 + (9l + 2m - 5\lambda - 6\mu) I_1{}^2 + (3m + n + 8\mu) I_2$$

(the invariant $\zeta_a{}^a$ of the reciprocal of the strain tensor $\epsilon_s{}^r$ being I_2/I_3). In the treatment by Seth (7) of the problem of finite strain the strain components were not truncated, as in the classical infinitesimal theory, by the omission of the second order terms, but the equations

$$T_s{}^r = \lambda I_1 \delta_s{}^r + 2\mu\epsilon_s{}^r; \qquad T = (3\lambda + 2\mu) I_1$$

of the infinitesimal theory were taken over with the following explanatory remark "Since this is the simplest tensor form that we can take, it is quite natural for us to assume that the stress-strain relations are governed by equations of the above type." From the discussion given above it is clear that simplicity is not a sufficiently compelling reason; for the whole strength obtained by a willingness to keep the second order terms in the strain components is sacrificed by the omission of second order terms—such as those that occur in the terms $-\lambda I_1{}^2\delta_s{}^r$ etc.,—in the expressions for the stress components.

4. The case of hydrostatic pressure; comparison of theory with experiment. In this simplest case the strain tensor is scalar $\epsilon_s{}^r = \epsilon\delta_s{}^r$ and $2\epsilon = 1 - (V_0/V)^{2/3}$. In most cases the stress is a pressure rather than a tension and ϵ is negative and so we write $T_s{}^r = -p\delta_s{}^r$; $\epsilon = -f$. Then $I_1 = -3f$; $I_2 = 3f^2$; $I_3 = -f^3$, and the formula connecting pressure with change of volume is

$$p = af + bf^2; \qquad f = \tfrac{1}{2}\{(V_0/V)^{2/3} - 1\}$$
$$a = 3\lambda + 2\mu; \qquad b = 15\lambda + 10\mu - 27l - 9m - n.$$

Observe that if we avail ourselves only of the two elastic constants λ, μ of the classical (infinitesimal) theory $b = 15\lambda + 10\mu = 5a$ so that our formula is a one constant one

$$p = a(f + 5f^2); \qquad f = \tfrac{1}{2}\{(V_0/V)^{2/3} - 1\}; \qquad a = 3\lambda + 2\mu.$$

For small compressions

$$f = \tfrac{1}{2}\{(V_0/V)^{2/3} - 1\} = \tfrac{1}{2}\{(1 - \Delta V/V_0)^{-(2/3)} - 1\}$$
$$= \tfrac{1}{3}(\Delta V/V_0) + \cdots; \qquad \Delta V = V_0 - V$$

and so the modulus of compression: $p \div \Delta V/V_0 = p \div 3f$; hence the modulus of compression is $a/3$. The following table gives a comparison of the theory with the experimental results of Bridgman (**10**) upon the compressibility of sodium. The quadratic formula $p = af + bf^2$ was used and the two constants at our disposal were determined by the experimental results at 2,000 atm. (the beginning of the experiment) and at 12,000 atm. (near the middle of the experimental range which ran from 2,000 to 20,000 atm., at intervals of 2,000 atm.). In this way the availability of the formula for purposes of *extrapolation* (12,000 to 20,000 atm.), as well as interpolation (2,000 to 12,000 atm.), was tested.

TABLE 1.

p	$\Delta V/V_0$	f	p (calculated)
2000	.0295	.0101	——
4000	.0552	.0193	4005
6000	.0779	.0278	6022
8000	.0981	.0356	8005
10000	.1165	.0430	10003
12000	.1332	.0500	——
14000	.1488	.0567	14008
16000	.1632	.0631	16014
18000	.1767	.0692	18006
20000	.1894	.0751	20007

$$a = 1.874 \times 10^5; \qquad b = 1.052 \times 10^6.$$

The largest discrepancy is that corresponding to a measured pressure of 6,000 atm., and a calculated pressure of 6,022 atm., an error of about one-third of 1%. The other calculated values are correct to within one-tenth of 1%, most being much closer. Attention should be called to the fact that $b = 5.6\ a$, so that a neglect of the third order terms in the expansion of $\rho_0\phi$ (which neglect would force b to be $5a$) would only disturb the agreement to the extent $.6af^2$. Thus the one constant formula $p = a(f + 5f^2)$ fits the data for sodium to an accuracy of within 1.5% over the range 2,000 atm., to 20,000 atm., the constant a having the value 1.92×10^5 (determined by the measurement at 12,000 atm.). The agreement is as good as could be expected since f is only

measured to an accuracy varying between .5% at 2,000 atm., to .07% at 20,000 atm.

A two constant formula does not give very good agreement with experimental results for liquids which are much more compressible than solids. However a three constant formula (which would result if the energy-density were expanded as far as fourth order terms) gives very good agreement. The following table gives the result of a comparison between the results of calculation from a three-constant formula $p = af + bf^2 + cf^3$ and the experimental results of Bridgman (**11**) on the compressibility of N-Amyl Iodide, at 0° temperature, under pressures varying between 500 atm., and 12,000 atm. Over this range the pressures calculated agreed with those measured to within less than one per cent.

TABLE 2.

N-Amyl Iodide (0°)

p (obs)	V/V_0	f	af	bf^2	cf^3	p (calc.)
500	.9685	.0108	444.9	49.4	10.2	———
1000	.9442	.0195	803.9	161.7	37.2	1002.8
1500	.9250	.0267	1099	302.3	134.7	1496.4
2000	.9094	.0327	1347	453.7	174.9	1975.6
3000	.8831	.0432	1780	792.9	404.2	2977.1
4000	.8624	.0518	2136	1142	698.7	3976.7
5000	.8451	.0593	2445	1496	1048	4989
6000	.8304	.0659	2717	1848	1438	———
7000	.8173	.0720	2967	2202	1871	7040
8000	.8064	.0771	3177	2525	2297	7999
9000	.7965	.0819	3375	2849	2753	8977
10000	.7873	.0864	3560	3171	3233	9964
11000	.7786	.0908	3741	3502	3752	10995
12000	.7706	.0948	3908	3822	4277	———

$$a = 512.04 \times 10^2; \qquad b = 424.8 \times 10^3; \qquad c = 501.2 \times 10^4.$$

In order to show the dependence of the coefficients a, b, c upon the temperature similar calculations were made from measurements at 50° C with the following result

$$a_{50} = 303.94 \times 10^2; \qquad b_{50} = 316.26 \times 10^3; \qquad c_{50} = 365.51 \times 10^4.$$

A similar three constant formula for sodium over the range 2,000 to 20,000 atmospheres (the constants being determined by the measurements at 2,000, 10,000 and 20,000 atmospheres) gave the values

$$a = 188.13 \times 10^3; \quad b = 101.1 \times 10^4; \quad c = 37.2 \times 10^4$$

and the correspondence between the observed and calculated values was:

p (obs)	4000,	6000,	8000,	12000,	14000,	16000,	18000
p (calc)	4008,	6014,	8003,	11979,	13984,	15976,	17984

an agreement to within one quarter of one per cent over the entire range.

5. The Young's modulus experiment. In this experiment the ends of a cylinder of length l are subjected to a uniform stress T, the sides being free from any applied force. The conditions of the problem are met by assuming $u = px$; $v = py$; $w = rz$ (where the z-axis is parallel to the generators of the cylinder) it being agreed that mass forces, such as the weight of the mass elements of the cylinder, may be neglected. The strain tensor is diagonal with diagonal elements

$$\epsilon_1{}^1 = \epsilon_2{}^2 = p - \tfrac{1}{2}p^2; \qquad \epsilon_3{}^3 = r - \tfrac{1}{2}r^2$$

and the stress tensor is consequently also diagonal. The numbers p, r must be such that $T_x{}^x = T_y{}^y = 0$; $T_z{}^z = T$. If $e = w/c$ is the relative extension we have, since $z = c + w$, $e = r(1 + e)$ so that $r = e - e^2 + \cdots$ implying

$$\epsilon_3{}^3 = e - (3/2)e^2 + \cdots; \qquad e = \epsilon_3{}^3 + (3/2)(\epsilon_3{}^3)^2 + \cdots.$$

For simplicity of notation we write ϵ for $\epsilon_3{}^3$ and set $\epsilon_1{}^1 = \epsilon_2{}^2 = -\sigma\epsilon_3{}^3$. If $f = -u/a = -v/b$ denotes, for the moment, the relative contraction, in a direction perpendicular to the applied stress, we have

$$\epsilon_1{}^1 = -f - \frac{3}{2}f^2 + \cdots \text{ so that } \sigma = (f + \frac{3}{2}f^2 + \cdots)/(e - \frac{3}{2}e^2 + \cdots)$$
$$= \frac{f}{e}\{1 + \frac{3}{2}(f + e) + \cdots\}.$$

To a first approximation $\sigma = f/e$ measures the ratio of the relative contraction, perpendicular to the applied stress, to the relative extension in the direction of the applied stress; we shall refer to σ as Poisson's ratio.

Since $T_x{}^x = T_y{}^y = 0$ we have $T = T_z{}^z = T_a{}^a$ and so

$$T = (3\lambda + 2\mu)I_1 + (9l + 2m - 5\lambda - 6\mu)I_1{}^2 + (3m + n + 8\mu)I_2.$$

We shall content ourselves with examining what partial explanation of the phenomena observed in the Young's modulus experiment may be obtained by

using only the two elastic constants λ, μ of the infinitesimal theory. On setting $l = 0$, $m = 0$, $n = 0$ we find

$$T = (3\lambda + 2\mu)I_1 - (5\lambda + 6\mu)I_1^2 + 8\mu I_2.$$

On putting in the values $I_1 = (1 - 2\sigma)\epsilon$, $I_2 = (\sigma^2 - 2\sigma)\epsilon^2$ we find

$$T = (3\lambda + 2\mu)(1 - 2\sigma)\epsilon - \{(5\lambda + 6\mu)(1 - 2\sigma)^2 + 8\mu(2\sigma - \sigma^2)\}\epsilon^2.$$

The relation $T_x^x = 0$ yields

$$\lambda - 2\sigma(\lambda + \mu) - \{8(\lambda + \mu)\sigma^2 - 2(3\lambda + \mu)\sigma + \lambda\}\epsilon = 0$$

or

$$\left\{\sigma - \frac{\lambda}{2(\lambda + \mu)}\right\}\{1 + (4\sigma - 1)\epsilon\} = 0.$$

Since $V_0/V = (1 + 2\sigma\epsilon)\sqrt{1 - 2\epsilon}$ the maximum value of ϵ is .5 and so, granting $\sigma > 0$, $1 + (4\sigma - 1)\epsilon > .5$ so that $\sigma = \lambda/2(\lambda + \mu)$. It is important to notice that this constancy of Poisson's ratio σ is not a mere approximation but an exact result. On inserting this value of σ in the expression for T we find

$$T = E\epsilon\left\{1 - \frac{2\lambda + 3\mu}{\lambda + \mu}\epsilon\right\}$$

where $E = \dfrac{(3\lambda + 2\mu)\mu}{\lambda + \mu}$ is Young's modulus. The first approximation, which would be furnished by the infinitesimal theory, is $T = E\epsilon$ but the significance of the quadratic formula of the finite theory is that the graph of T against ϵ is a parabola instead of a straight line. Hence T has a maximum value $(T)_{\max} = \dfrac{\lambda + \mu}{4(2\lambda + 3\mu)}E$ occurring when $\epsilon = \dfrac{\lambda + \mu}{2(2\lambda + 3\mu)}$. What this means is that if a larger stress is applied the deformation cannot be of the simple type described by the formulae $u = px$; $v = py$; $w = rz$. For steel λ is approximately 1.5μ so that $(T)_{\max} = E/10$ the corresponding value of ϵ being .2. In using the formula given above for T it should be noticed that $\epsilon = r - \frac{1}{2}r^2 = \dfrac{(2 + e)e}{2(1 + e)^2}$ where e is the relative extension. These results of the finite theory, predicting a maximum stress (yield point) are qualitatively correct but the predicted value of $(T)_{\max}$ is much too large.

6. The stress-strain equations for a non-isotropic medium. The elastic potential is now a function $\phi({}_{pq}\eta, {}_{pq}c, T)$ of the scalar strain components ${}_{pq}\eta$, the coefficients ${}_{pq}c$ of the quadratic form for $(ds_0)^2$ and the temperature T. On using the expression for $\delta\,{}_{pq}\eta$ given in 1, namely

$$2\delta\,{}_{pq}\eta = \{(\delta x_\alpha)_{,\beta} + (\delta x_\beta)_{;\alpha}\}({}_{p,}x^\alpha)({}_{q,}x^\beta)$$

and agreeing that ϕ is symmetrized with respect to the scalar strain components ${}_{pq}\eta$ so that $\partial\phi/\partial({}_{pq}\eta) = \partial\phi/\partial({}_{qp}\eta)$ we find at once

$$T^{rs} = \rho\,\frac{\partial\phi}{\partial({}_{\alpha\beta}\eta)}\,{}_{\alpha,}x^r\,{}_{\beta,}x^s.$$

These expressions give, from the Lagrangian viewpoint, the stress tensor in terms of the gradient of the elastic potential relative to the scalar strain components ${}_{pq}\eta$. In order to obtain the corresponding Eulerian equations we introduce the matrix $\boldsymbol{j}$ reciprocal to the matrix $\boldsymbol{k}$ whose elements

$${}_{pq}k = g_{\alpha\beta}({}_{p,}x^\alpha)({}_{q,}x^\beta) = 2\,{}_{pq}\eta + {}_{pq}c\,;$$

it being clear that the elements of $\boldsymbol{j}$ are given by the formula

$${}^{pq}j = g^{\alpha\beta}({}^{p}a_{,\alpha})({}^{q}a_{,\beta}).$$

On taking the variation of the matrix equation $\boldsymbol{jk} = \boldsymbol{e}$ (the unit matrix) we find $\delta\boldsymbol{j}\cdot\boldsymbol{k} + \boldsymbol{j}\cdot\delta\boldsymbol{k} = \boldsymbol{0}$ or equivalently, $\delta\boldsymbol{j} = -\boldsymbol{j}\cdot\delta\boldsymbol{k}\cdot\boldsymbol{j}$. Hence

$$\begin{aligned}\frac{\partial\phi}{\partial({}_{\alpha\beta}\eta)}\,\delta({}_{\alpha\beta}\eta) = \frac{\partial\phi}{\partial({}^{\alpha\beta}j)}\,\delta({}^{\alpha\beta}j) &= -\frac{\partial\phi}{\partial({}^{\alpha\beta}j)}\,({}^{\alpha\sigma}j)\,\delta({}_{\sigma\tau}k)\,({}^{\tau\beta}j)\\ &= -2\,\frac{\partial\phi}{\partial({}^{\alpha\beta}j)}\,({}^{\alpha\sigma}j)\,\delta({}_{\sigma\tau}\eta)\,({}^{\tau\beta}j)\end{aligned}$$

so that

$$\frac{\partial\phi}{\partial({}_{pq}\eta)} = -2\,({}^{p\alpha}j)\,\frac{\partial\phi}{\partial({}^{\alpha\beta}j)}\,({}^{\beta q}j).$$

Since

$$({}^{p\alpha}j)({}_{\alpha,}x^r) = g^{r\alpha}({}^{p}a_{,\alpha}) = {}^{p}a^{,r}$$

it follows that

$$T^{rs} = -2\rho\,\frac{\partial\phi}{\partial({}^{\alpha\beta}j)}\,({}^{\alpha}a^{,r})({}^{\beta}a^{,s}).$$

Appendix.

Expression of the quantities J_1, J_2, J_3 in terms of the invariants I_1, I_2, I_3. The invariants s_1, s_2, s_3 of any matrix $\boldsymbol{u}$ are determined by the equation

$$\det(\lambda\delta_s{}^r - u_s{}^r) = \lambda^3 - s_1\lambda^2 + s_2\lambda - s_3.$$

It is immediately clear that if $\boldsymbol{u}, \boldsymbol{v}$ are any two non-singular 3×3 matrices the two products $\boldsymbol{uv}$ and $\boldsymbol{vu}$ have the same invariants. In fact if $\boldsymbol{E} = (\delta_s{}^r)$ is the three-rowed unit matrix

$$\det(\lambda \boldsymbol{E} - \boldsymbol{vu}) = \det \boldsymbol{u}^{-1}(\lambda \boldsymbol{E} - \boldsymbol{uv})\boldsymbol{u} = \det(\lambda \boldsymbol{E} - \boldsymbol{uv}).$$

Secondly if $\boldsymbol{u}$ is any non-singular three-rowed matrix with invariants s_1, s_2, s_3 the invariants $\sigma_1, \sigma_2, \sigma_3$ of its reciprocal $\boldsymbol{u}^{-1}$ are given by the formulae $\sigma_1 = s_2/s_3$; $\sigma_2 = s_1/s_3$; $\sigma_3 = 1/s_3$. For

$$\det(\lambda \boldsymbol{E} - \boldsymbol{u}^{-1}) = -\lambda^3 (\det \boldsymbol{u})^{-1} \det\left(\frac{1}{\lambda}\boldsymbol{E} - \boldsymbol{u}\right) = \frac{1}{s_3}\{\lambda^3 s_3 - \lambda^2 s_2 + \lambda s_1 - 1\}$$

since $\det \boldsymbol{u} = s_3$.

Now if we denote by $\boldsymbol{g}$ the matrix whose elements are g_{pq} and by $\boldsymbol{t}$ the matrix whose elements are ${}_{s,}x^r$ (the upper label denoting the row and the lower the column) the matrix whose elements are ${}_{pq}k = g_{\alpha\beta}({}_{p,}x^\alpha)({}_{q,}x^\beta)$ is $\boldsymbol{t'gt}$ where the prime attached to a matrix denotes, as usual, its transposed. Hence the matrix $\boldsymbol{k}$ whose elements are

$${}^p{}_q k = ({}^{p\alpha}c)({}_{\alpha q}k) = 2({}^p{}_q\eta) + {}^p{}_q\delta$$

is $\boldsymbol{c}^{-1}\boldsymbol{t'gt}$ ($\boldsymbol{c}$ being the matrix whose elements are ${}_{pq}c$). Similarly the matrix whose elements are $h_{pq} = {}_{\alpha\beta}c({}^\alpha a_{,p})({}^\beta a_{,q})$ is $\boldsymbol{t}^{-1\prime}\boldsymbol{ct}^{-1}$ since the matrix whose elements are ${}^r a_{,s}$ is the reciprocal of the matrix whose elements are ${}_{s,}x^r$. Hence the matrix $\boldsymbol{h}$ whose elements are $h_q{}^p = \delta_q{}^p - 2\epsilon_q{}^p$ is $\boldsymbol{g}^{-1}(\boldsymbol{t}^{-1})'\boldsymbol{ct}^{-1}$. Hence its reciprocal $\boldsymbol{h}^{-1}$ is given by $\boldsymbol{h}^{-1} = \boldsymbol{tc}^{-1}\boldsymbol{t'g}$ and since $\boldsymbol{k} = \boldsymbol{c}^{-1}\boldsymbol{t'gt}$ it follows that $\boldsymbol{k}$ and $\boldsymbol{h}^{-1}$ have the same invariants. But the invariants of $\boldsymbol{k}$ are the coefficients of $-\lambda^2, \lambda, -1$, respectively, in the development of

$$\begin{aligned}\det(\lambda \boldsymbol{E} - \boldsymbol{k}) &= \det[(\lambda - 1)\boldsymbol{E} - 2\boldsymbol{\eta}] \\ &= 8\det(\mu \boldsymbol{E} - \boldsymbol{\eta}); \qquad \mu = \frac{\lambda - 1}{2} \\ &= (\lambda - 1)^3 - 2(\lambda - 1)^2 J_1 + 4(\lambda - 1)J_2 - 8J_3.\end{aligned}$$

Hence the invariants of $\boldsymbol{k}$ are, respectively,

$$2J_1 + 3, \qquad 4J_2 + 4J_1 + 3, \qquad 8J_3 + 4J_2 + 2J_1 + 1.$$

Similarly the invariants of $\boldsymbol{h}$ are, respectively,

$$3 - 2I_1, \qquad 3 - 4I_1 + 4I_2, \qquad 1 - 2I_1 + 4I_2 - 8I_3$$

so that the invariants of $\boldsymbol{h}^{-1}$ are, respectively,

$$\frac{3-4I_1+4I_2}{1-2I_1+4I_2-8I_3}; \qquad \frac{3-2I_1}{1-2I_1+4I_2-8I_3}; \qquad \frac{1}{1-2I_1+4I_2-8I_3}.$$

Hence we have the equations

$$2J_1+3=\frac{3-4I_1+4I_2}{1-2I_1+4I_2-8I_3}; \quad 4J_2+4J_1+3=\frac{3-2I_1}{1-2I_1+4I_2-8I_3},$$

$$8J_3+4J_2+2J_1+1=\frac{1}{1-2I_1+4I_2-8I_3}$$

and solving these we find

$$J_1=\frac{I_1-4I_2+12I_3}{1-2I_1+4I_2-8I_3}; \; J_2=\frac{I_2-6I_3}{1-2I_1+4I_2-8I_3}; \; J_3=\frac{I_3}{1-2I_1+4I_2-8I_3}$$

or, equivalently,

$$I_1=\frac{J_1+4J_2+12J_3}{1+2J_1+4J_2+8J_3}; \; I_2=\frac{J_2+6J_3}{1+2J_1+4J_2+8J_3}; \; I_3=\frac{J_3}{1+2J_1+4J_2+8J_3}.$$

Summary.

In the present paper formulae are derived which enable one to calculate the stress in an elastic medium when the strain and the elastic energy density are known, no simplifying assumptions, such as smallness of strain, being necessary. For an isotropic elastic solid under hydrostatic pressure the following *one* constant formula gives good agreement with experimental observation (only two elastic constants λ, μ being used in the expression for the elastic energy density)

$$p=a(f+5f^2); \qquad f=\tfrac{1}{2}\{(V_0/V)^{2/3}-1\}; \qquad a=3\lambda+2\mu.$$

In the Young's modulus experiment the formula for the extensional stress (again using only the two constants λ, μ) is

$$T=E\epsilon\left\{1-\frac{2\lambda+3\mu}{\lambda+\mu}\epsilon\right\}; \qquad E \text{ (Young's modulus)} = \frac{(3\lambda+2\mu)\mu}{\lambda+\mu}$$

where $\epsilon=\dfrac{(2+e)e}{(1+e)^2}$, e being the relative extension. Hence T has a maximum value $\dfrac{\lambda+\mu}{4(2\lambda+3\mu)}E$, occurring when $\epsilon=\dfrac{\lambda+\mu}{2(2\lambda+3\mu)}$. For a true second

order approximation (the infinitesimal theory being regarded as a first order approximation) five elastic constants occur and the corresponding formulae are either given or their derivation is immediate.

THE JOHNS HOPKINS UNIVERSITY;
INSTITUTE FOR ADVANCED STUDY.

REFERENCES.

1. G. Kirchhoff, "Über die Gleichungen des Gleichgewichtes eines elastischen Körpers bei nicht unendlich kleinen Verschiebungen seiner Theile," *Sitz. math-nat. Klasse der kaiserlichen Akad. der Wiss.*, **9** (1852), p. 762.
2. J. Boussinesq, "Théorie des ondes liquides périodiques," *Mémoires présentés à l'Ac. des Sciences,* **20** (1869), p. 516.
3. E. et F. Cosserat, "Sur la théorie de l'élasticité," *Ann. de Toulouse,* **10** (1896).
4. P. Duhem, "Recherches sur l'Élasticité," *Ann. de l'École nor. sup.*, 3me Série, **21, 22, 23** (1904-06).
5. E. Almansi, "Sulle deformazioni finite dei solidi elastici isotropi," *Rend. Lincei,* Ser. 5, Tome **20**[1] (1911).
6. L. Brillouin, "Sur les tensions de radiation," *Ann. de Physique,* 10me Série 4 (1925).
7. B. R. Seth, "Finite strain in elastic problems," *Phil. Trans. Roy. Soc.,* A **234** (1935), p. 231.
8. A. Signorini, "Trasformazioni termoelastiche finite, etc.," *Soc. Italiana per il progresso delle scienze* **14** (1936).
9. E. G. Coker, and L. N. G. Filon, *Treatise on photo-elasticity,* Cambridge (1931).
10. P. W. Bridgman, "Electrical resistances and volume changes up to 20,000Kg./cm.^{2}," *Proceedings of the National Academy of Sciences,* **21** (1935), p. 109.
11. P. W. Bridgman, "The pressure-volume-temperature relations of fifteen liquids," *Proc. Amer. Ac. Arts and Sc.,* **68** (1933).

MEAN MOTIONS AND DISTRIBUTION FUNCTIONS.[1]

By Philip Hartman, E. R. van Kampen and Aurel Wintner.

Let $k = 1, \cdots, n$ and $\Sigma = \sum_{k=1}^{n}$. If $z(t)$ is a function of the form

$$z(t) \equiv x(t) + iy(t) = \Sigma\, a_k \exp 2\pi i(\lambda_k t + \alpha_k), \tag{1}$$

where λ_k, α_k are real and $a_k > 0$, put

$$z(t) = \pm \mid z(t) \mid \exp 2\pi i \phi(t), \tag{2}$$

where the sign of $\pm \mid z(t) \mid$ is to be chosen for every t in such a way that $\phi = \phi(t)$ becomes a continuous function of t. The function $\phi(t)$ is said to have for $t \rightarrow +\infty$ a mean motion μ if

$$\phi(t)/t \rightarrow \mu, \text{ i. e., } \phi(t) = \mu t + o(t);\ t \rightarrow +\infty. \tag{3}$$

The problem of the existence and the determination of this constant μ goes back to Lagrange's approximative treatment of secular perturbations of the major planets and has been solved in the case $n = 3$ by Bohl.[2] The case $n = 4$ has been treated by Weyl.[3] The present note attempts a general approach to the problem from the point of view of the theory of distribution functions.

The connection between mean motions and asymptotic distribution functions is suggested by the following consideration: It is known [4] that if the real function $\xi = \psi(t)$ is almost periodic in the sense of Bohr, then $\psi(t)$ possesses an asymptotic distribution function $\sigma(\xi)$ and one has, for every continuous function $f = f(\xi)$,

$$\int_{-\infty}^{+\infty} f(\xi)\, d\sigma(\xi) = M\{f(\psi(t))\}, \tag{4}$$

where

$$M\{g(t)\} = \lim_{T \rightarrow +\infty} (1/T) \int_0^T g(t)\, dt. \tag{5}$$

It is understood that by the existence of an asymptotic distribution function of a real measurable function $\psi(t)$, where $0 \leqq t < +\infty$, is meant the existence of a monotone function $\sigma(\xi)$ such that $\sigma(-\infty) = 0$, $\sigma(+\infty) = 1$ and, if ξ is a continuity point of $\sigma(\xi)$,

[1] Received January 20, 1937.

[2] Bohl [3].

[3] Weyl [7].

[4] Wintner [8].

$$(1/T) \text{ meas } [\psi(t) \leqq \xi]_T \to \sigma(\xi), \qquad T \to +\infty,$$

where $[\psi(t) \leqq \xi]_T$ denotes the set of those points t of the interval $0 \leqq t \leqq T$ at which $\psi(t) \leqq \xi$. Now suppose that the amplitudes a_k of (1) satisfy the condition of Lagrange, i. e., that

$$a_{k_1} > a_{k_2} + \cdot \cdot \cdot + a_{k_n} \tag{6}$$

holds for a permutation $(k_1, \cdot \cdot \cdot, k_n)$ of $(1, \cdot \cdot \cdot, n)$. Then (3) may be replaced by the sharper statement that

$$\phi(t) = \mu t + \omega(t), \tag{3 bis}$$

where $\omega(t)$ and also its derivative $\omega'(t)$ are almost periodic in the sense of Bohr.[5] Hence, on denoting by $\psi(t)$ the almost periodic function $\phi'(t) = \mu + \omega'(t)$ and by $\sigma(\xi)$ its asymptotic distribution function, (4) is applicable, and becomes, when $f(\xi) = \xi$,

$$\int_{-\infty}^{+\infty} \xi d\sigma(\xi) = M\{\phi'(t)\}, \tag{7}$$

where, according to (5) and (3),

$$M\{\phi'(t)\} = \lim_{T \to +\infty} \phi(T)/T = \mu. \tag{8}$$

Thus

$$\int_{-\infty}^{+\infty} \xi d\sigma(\xi) = \mu, \tag{9}$$

so that the mean motion of $\phi(t)$ appears as the first moment of the asymptotic distribution function $\sigma(\xi)$ of $\phi'(t)$.

If the amplitudes a_k do not satisfy the inequality (6), the preceding considerations break down in view of the fact that $\phi'(t)$ is not, in general, almost periodic in the sense of Bohr, while a possible treatment of the problem within a class of almost periodic functions more general than those of Bohr leads to difficulties. In fact, one needs the particular case (7) of (4), and (7) is obvious only when ψ $(= \phi')$ is a bounded function, a condition which is not satisfied in the majority of cases, if (6) does not hold. It is not difficult to prove that the function $\phi'(t)$ has a distribution function $\sigma(\xi)$ and that the space-average represented by the Stieltjes integral on the left of (7) is absolutely convergent. The main difficulty arises in the identification of the

[5] Wintner [9].

space-average with the corresponding time-average $M\{\phi'(t)\}$. It is known [6] that the truth of Lindelöf's hypothesis in the theory of the Riemann zeta-function depends on a question of the same type, namely on the question as to the admissibility of the identification of certain space-averages with the corresponding (hypothetical) time-averages, as expressed by (4).

It will be assumed that the frequencies λ_k of (1) are linearly independent. This assumption, as will be seen from the proof, does not essentially affect the validity of the method. After proving the existence of the asymptotic distribution function $\sigma(\xi)$ and of its first moment, the identification of the time-average with the space-average remains to be treated. The admissibility of this identification will be proved with the help of Birkhoff's ergodic theorem,[7] by excluding, for fixed values of the amplitudes a_k and the frequencies λ_k, a set of measure zero in the n-dimensional space of the phases α_k. Actually, there are some indications that Birkhoff's zero set is empty in the present case. The assumption of the linear independence of the frequencies λ_k is to the effect [8] that the problem is of the metrically transitive type, so that the mean motion μ will depend on the amplitudes a_k and the frequencies λ_k but not on the phases α_k.

The explicit evaluation of μ will be reduced for every n to the evaluation of a definite integral in the α_k-space. On comparing the results of the present paper with those of Bohl [2] $(n=3)$, it follows that if a_1, a_2, a_3 are positive, then the integral

$$\int_0^1\int_0^1 \frac{[a_1{}^2+a_1a_2\cos 2\pi(\vartheta_2-\vartheta_1)+a_1a_3\cos 2\pi(\vartheta_3-\vartheta_1)]\,d\vartheta_1\,d\vartheta_2\,d\vartheta_3}{a_1{}^2+a_2{}^2+a_3{}^2+2a_1a_2\cos 2\pi(\vartheta_2-\vartheta_1)+2a_1a_3\cos 2\pi(\vartheta_3-\vartheta_1)+2a_2a_3\cos 2\pi(\vartheta_3-\vartheta_2)}$$

$$\int_0^1\int_0^1 \frac{[a_1{}^2+a_1a_2\cos 2\pi\vartheta_1+a_1a_3\cos 2\pi\vartheta_2]\,d\vartheta_1\,d\vartheta_2}{a_1{}^2+a_2{}^2+a_3{}^2+2a_1a_2\cos 2\pi\vartheta_1+2a_1a_3\cos 2\pi\vartheta_2+2a_2a_3\cos 2\pi(\vartheta_1-\vartheta_2)}$$

is equal to

$$\frac{1}{\pi}\operatorname{arc}\cos(a_2{}^2+a_3{}^2-a_1{}^2)/(2a_2a_3)$$

if $a_i \leqq a_j+a_k$ for all permutations (i,j,k) of $(1,2,3)$, while it is equal to 1 if $a_1>a_2+a_3$, and finally it is equal to 0 if either $a_2>a_1+a_3$ or $a_3>a_1+a_2$. Similar relations follow by comparison of the results of the present paper with those of Weyl [3] $(n=4)$. These definite integrals show with varying a_k "discontinuities" of the same type as the well-known "discontinuous" integrals

[6] Cf., on the one hand, Hardy and Littlewood [4] and, on the other hand, Jessen and Wintner [5], Theorem 31.

[7] Birkhoff [1]; cf. also Khintchine [6].

[8] Cf. Birkhoff [2], p. 371.

of Sonine ($n=3$) and Nicholson ($n=4$), occurring in connection with Lord Rayleigh's random walk problem and presented on pp. 411, 414 and 420 of Watson's Treatise on Bessel functions.

For a given function (1), put

$$\psi(t)=\tfrac{1}{2\pi}\frac{x(t)y'(t)-y(t)x'(t)}{[x(t)]^2+[y(t)]^2},\ \text{if } z(t)\neq 0, \tag{10}$$

where the prime denotes differentiation with respect to t. It is clear that, no matter which locally continuous determination is chosen for

$$\arg z(t) = -i\log[z(t)/|z(t)|],$$

one has

$$[\arg z(t)]' = 2\pi\psi(t),\ \text{if } z(t)\neq 0.$$

It is known [2] that a unique function $\phi(t)$ is defined by the following requirements:

(i) $\phi(t)$ is continuous for $-\infty<t<+\infty$;
(ii) $\phi'(t)=\psi(t)$, if $z(t)\neq 0$;
(iii) $2\pi\phi(t)\equiv\arg z(t) \pmod{\pi}$, if $z(t)\neq 0$;
(iv) $0\leqq\phi(0)<\frac{1}{2}$.

The function $\phi(t)$ thus defined satisfies (2), where one has to choose, for every t, the sign $+$ or the sign $-$ according as the difference

$$|2\pi\phi(t)-\arg z(t)|$$

is an even or an odd multiple of π, while there is no ambiguity in the case $z(t)=0$. The function $\phi(t)$ will be referred to as the angular function belonging to (1). The values of t for which $z(t)=0$, i. e., for which $\phi'(t)=\psi(t)$ is undefined, clearly do not have a cluster point.

Let $\chi=\chi(\vartheta_1,\cdots,\vartheta_n)$ denote the function

$$\chi=\frac{(\Sigma\lambda_k a_k\cos 2\pi\vartheta_k)(\Sigma a_k\cos 2\pi\vartheta_k)+(\Sigma\lambda_k a_k\sin 2\pi\vartheta_k)(\Sigma a_k\sin 2\pi\vartheta_k)}{(\Sigma a_k\cos 2\pi\vartheta_k)^2+(\Sigma a_k\sin 2\pi\vartheta_k)^2} \tag{11}$$

defined on the n-dimensional torus

$$\Theta:\quad 0\leqq\vartheta_k<1; \qquad (k=1,\cdots,n), \tag{12}$$

except on the set N of those points of Θ at which the denominator of (11) vanishes, so that N is the set on Θ defined by the pair of equations

$$\mathrm{N}:\quad F\equiv\Sigma a_k\cos 2\pi\vartheta_k=0, \qquad G\equiv\Sigma a_k\sin 2\pi\vartheta_k=0. \tag{13}$$

Suppose that the frequencies λ_k of (1) are linearly independent and let

Z denote the curve on the torus (12) which is defined by the parameter representation

(14) $$Z:\ \vartheta_k = \lambda_k t + \alpha_k \pmod 1; \qquad (k = 1, \cdots, n),$$

where the parameter t runs from 0 to $+\infty$ and the phases α_k are arbitrarily fixed.

First, it will be shown that the derivative $\phi'(t)$ of the angular function $\phi(t)$ of the function (1) has an asymptotic distribution function $\sigma(\xi)$ and that

(15) $$\sigma(\xi) = \text{meas}\, \Gamma_\xi, \qquad -\infty < \xi < +\infty,$$

where the meas Γ_ξ denotes the n-dimensional $(\vartheta_1, \cdots, \vartheta_n)$-measure of the set Γ_ξ of those points $(\vartheta_1, \cdots, \vartheta_n)$ of the torus (12) at which the function (11) satisfies the inequality

(16) $$\chi \leqq \xi.$$

In order to prove this, notice first that the set (13) at which the denominator of (11) vanishes clearly has a vanishing Jordan content (a detailed description of N will be given below). It is also clear from (11) that Γ_ξ has for every ξ a Jordan content. On the other hand, the linear independence of the λ_k implies that to distinct values of t there belong distinct points of the curve (14) on the torus (12). Furthermore, it is seen from (1), (10) and (11) that

(17) $$\chi(\lambda_1 t + \alpha_1, \cdots, \lambda_n t + \alpha_n) = \psi(t) = \phi'(t)$$

holds at those points of the curve (14) which do not lie on the subset (13) of the torus (12), i. e., at which $z(t) \neq 0$. Now it is clear from (17) and from the definition of Γ_ξ that $\psi(t) \leqq \xi$ holds if and only if that point of the curve (14) to which t belongs is a point of Γ_ξ. It follows, therefore, from the Kronecker-Weyl approximation theorem that if $\{\xi; T\}$ denotes the sum of the lengths of those subintervals of the interval $0 \leqq t \leqq T$ on which $\psi(t) \leqq \xi$, then

$$\{\xi; T\}/T \to \text{meas}\, \Gamma_\xi, \qquad T \to +\infty.$$

This proves that $\psi(t) = \phi'(t)$ has the function (15) as asymptotic distribution function.

Next, it will be shown that the set N defined by (13) is a closed, possibly disconnected, $(n-2)$-dimensional analytic manifold in the n-dimensional torus $\mathfrak{S}$ defined by (12), and that the manifold N has no singularities or a finite number of singular curves according as there does not or does exist at least one permutation $(i_1, \cdots, i_n)$ of $(1, \cdots, n)$ such that

(18) $$a_{i_1} + \cdots + a_{i_m} = a_{i_{m+1}} + \cdots + a_{i_n}$$

holds for some m. It will be seen from the proof that if $n=2$ or $n=3$, then N consists of at most two analytic simple closed curves without singularities. Incidentally, N is, for an arbitrary n, empty if and only if the condition (6) of Lagrange is satisfied. This is clear from the definition (13) of N.

In order to prove that N has the structure described above, let j, l be a pair of distinct values of $k=1,\cdots,n$ and let J_{jl} denote the Jacobian with respect to ϑ_j, ϑ_l of the two functions F, G occurring in the definition (13) of N, so that

$$(19)\qquad J_{jl}=\frac{\partial(F,G)}{\partial(\vartheta_j,\vartheta_l)}=4\pi^2 a_j a_l \sin 2\pi(\vartheta_l-\vartheta_j)\,;\qquad j\neq l.$$

Accordingly, a point $\mathrm{P}=(\vartheta_1,\cdots,\vartheta_n)$ of the subset N of the torus Θ is a singular point of the manifold N if and only if

$$(20)\qquad 2\,|\,\vartheta_j-\vartheta_l\,|\equiv 0 \pmod 1,\quad \text{where}\quad j,l=1,\cdots,n;\ j\neq l.$$

Suppose that there exists on N a point $\mathrm{P}=(\vartheta_1,\cdots,\vartheta_n)$ satisfying (20). Then one can arrange the numbers $\vartheta_1,\cdots,\vartheta_n$ into two groups

$$(\vartheta_{i_1},\cdots,\vartheta_{i_m}),\qquad(\vartheta_{i_{m+1}},\cdots,\vartheta_{i_n})$$

such that

$$(21\text{a})\qquad \vartheta_{i_r}-\vartheta_{i_s}=\tfrac{1}{2}\,;\qquad (r=1,\cdots,m;\ s=m+1,\cdots,n),$$

while

$$(21\text{b})\qquad \vartheta_{i_1}=\cdots=\vartheta_{i_m},\qquad \vartheta_{i_{m+1}}=\cdots=\vartheta_{i_n}.$$

Now it is clear that the pair of conditions (21a), (21b) defines on the torus Θ a simple closed curve, and that this simple closed curve lies, in view of (20) and (13), on the $(n-2)$-dimensional manifold N if and only if the amplitudes a_k satisfy (18).

In what follows, it will be assumed for the sake of brevity that the amplitudes a_k of (1) do not satisfy an equation of the form (18), i. e., that the manifold N is free of singularities, so that no point of Θ which satisfies (20) is a point of N.

In order to prove that the space-average occurring on the left-hand side of (7) is finite for the distribution function (15), one has, according to the definition of Γ_ξ, merely to prove that the function (11) is absolutely integrable over the torus Θ. Actually,

$$(22)\qquad \int_{\Theta}|\,\chi\,|^{\kappa}\,d\Theta=\int_0^1\cdots\int_0^1|\,\chi(\vartheta_1,\cdots,\vartheta_n)\,|^{\kappa}\,d\vartheta_1\cdots d\vartheta_n<+\infty,\ \text{if}\ 0<\kappa<2.$$

The proof of (22) proceeds as follows: Let

$$(23)\qquad \mathrm{P}^0\colon\ \vartheta_k=\vartheta_k{}^0\,;\qquad (k=1,\cdots,n),$$

be a point of the manifold N, so that, since N is free of singularities, (20) is not satisfied by $(\vartheta_1, \cdots, \vartheta_n) = (\vartheta_1^0, \cdots, \vartheta_n^0)$, and so there exists at least one pair j, l, say $j = 1$ and $l = 2$, such that

(24) $$2 \mid \vartheta_1^0 - \vartheta_2^0 \mid \not\equiv (\operatorname{mod} 1), \text{ i. e., } \cos 2\pi(\vartheta_1^0 - \vartheta_2^0) \neq \pm 1.$$

Now the quadratic form

(25) $$Q(u, v; \vartheta_1, \vartheta_2) = 4\pi^2[a_1^2u^2 + 2a_1a_2 \cos 2\pi(\vartheta_1 - \vartheta_2)uv + a_2^2v^2]$$

is such that for a sufficiently small $\epsilon > 0$ and for some positive $\eta = \eta(\epsilon; \mathrm{P}^0)$ one has

(26) $$Q(u, v; \vartheta_1, \vartheta_2) \geqq (u^2 + v^2)\eta, \text{ if } \mid \vartheta_1 - \vartheta_1^0 \mid < \epsilon, \mid \vartheta_2 - \vartheta_2^0 \mid < \epsilon.$$

This follows for reasons of continuity from the fact that $Q(u, v; \vartheta_1^0, \vartheta_2^0)$ is positive definite in view of (24) and (25). On the other hand, if a point

(27) $$\mathrm{P}^* = (\vartheta_1^*, \cdots, \vartheta_n^*)$$

is on N, then it is seen from (13) and from (25) that

(28) $$[F(\vartheta_1^* + u, \vartheta_2^* + v, \vartheta_3^*, \cdots, \vartheta_n^*)]^2 + [G(\vartheta_1^* + u, \vartheta_2^* + v, \vartheta_3^*, \cdots, \vartheta_n^*)]^2$$
$$= Q(u, v; \vartheta_1^*, \vartheta_2^*) + o(u^2 + v^2) \text{ as } u^2 + v^2 \to 0,$$

and that the o-term holds uniformly for all choices of the point (27) on N. Hence it is seen from (26) that for a sufficiently small $\delta > 0$ and for some positive $\zeta = \zeta(\delta; \mathrm{P}^0)$ one has

(29) $$[F(\vartheta_1^* + u, \vartheta_2^* + v, \vartheta_3^*, \cdots, \vartheta_n^*)]^2$$
$$+ [G(\vartheta_1^* + u, \vartheta_2^* + v, \vartheta_3^*, \cdots, \vartheta_n^*)]^2 \geqq (u^2 + v^2)\zeta$$

whenever (27) is a point of N such that

(30) $$\mid \vartheta_1^* - \vartheta_1^0 \mid < \delta, \mid \vartheta_2^* - \vartheta_2^0 \mid < \delta, \text{ while } u^2 + v^2 < \delta^2.$$

Now it is clear from (11) and (13) that

(31) $$\mid \chi(\vartheta_1, \cdots, \vartheta_n) \mid \leqq \text{const.} \{[F(\vartheta_1, \cdots, \vartheta_n)]^2 + [G(\vartheta_1, \cdots, \vartheta_n)]^2\}^{-\frac{1}{2}}$$

for all points $(\vartheta_1, \cdots, \vartheta_n)$ of $\mathfrak{Q}$ which do not lie on N. It is seen from (29) and (31) that

(32) $$\mid \chi(\vartheta_1^* + u, \vartheta_2^* + v, \vartheta_3^*, \cdots, \vartheta_n^*) \mid = O([u^2 + v^2]^{-\frac{1}{2}}), \; u^2 + v^2 \to 0,$$

holds uniformly with respect to P*, if P^0 is fixed and (30) is satisfied. This clearly implies that the contribution of a sufficiently small vicinity of P^0 to the n-fold integral (22) is finite. Since P^0 is any point of the closed, bounded,

$(n-2)$-dimensional manifold (13), and since this manifold consists of the zeros of the denominator of (11), the proof of (22) is complete.

The admissibility of the identification of the space-average with the time-average as mentioned at the beginning of the paper may be treated as follows: Consider the transformation

$$\tau_t = \tau_t(\vartheta_1, \cdots, \vartheta_n)$$

which sends a point

$$(\vartheta_1, \cdots, \vartheta_n)$$

of Θ into the point

$$(\lambda_1 t + \vartheta_1, \cdots, \lambda_n t + \vartheta_n)$$

of Θ. Thus τ_t is a measure-preserving transformation of Θ into itself, satisfies the group condition $\tau_r \tau_s = \tau_{r+s}$ and is of the metrically transitive type [8] in view of the linear independence of the λ_k. Hence the ergodic theorem of Birkhoff [7] is applicable to every L-integrable function $\nu = \nu(\vartheta_1, \cdots, \vartheta_n)$ on Θ and thus, according to (22), to the function $\nu = \chi$. Hence on excluding from Θ a set of points $(\vartheta_1, \cdots, \vartheta_n)$ of measure zero, the time-average (3) of the function

$$g(t) = \chi(\lambda_1 t + \vartheta_1, \cdots, \lambda_n t + \vartheta_n)$$

exists and is equal to the integral of the function (11) over the torus Θ. This means in view of (15), (17) and (1) that on keeping the frequencies λ_k and the amplitudes a_k fixed and on excluding from the n-dimensional space of the phases α_k a set of measure zero, (7) holds for the derivative $\phi'(t)$ of the angular function $\phi(t)$ of (1) and for the asymptotic distribution function $\sigma(\xi)$. Finally, (3) follows from (7) in view of (8).

It may be mentioned that on excluding, for fixed a_k and λ_k, a set of phases $(\alpha_1, \cdots, \alpha_n)$ which is $(n-1)$-dimensional, hence of measure zero, the function (1) is distinct from zero for every t, so that one can choose in (2) the sign $+$ for every t, in which case $|z(t)|$ and $\phi(t)$ become polar coördinates in the (x, y)-plane. It is known [9] that, in virtue of the linear independence of the λ_k, the asymptotic distribution function of the polar *angle* $\phi(t)$ thus defined is a circular equidistribution, since the asymptotic distribution of (1) in the (x, y)-plane is of radial symmetry.

THE JOHNS HOPKINS UNIVERSITY.

BIBLIOGRAPHY.

[1] G. D. Birkhoff, "Proof of the ergodic theorem," *Proceedings of the National Academy of Sciences*, vol. 17 (1931), pp. 656-660.

[9] Wintner [10].

[2] G. D. Birkhoff, "Probability and physical systems," *Bulletin of the American Mathematical Society*, vol. 38 (1932), pp. 361-379.

[3] P. Bohl, "Über ein in der Theorie der säkularen Störungen vorkommendes Problem," *Journal für die reine und angewandte Mathematik*, vol. 135 (1909), pp. 189-283.

[4] G. H. Hardy and J. E. Littlewood, "On Lindelöf's hypothesis concerning the Riemann zeta-function," *Proceedings of the Royal Society*, ser. A, vol. 103 (1923), pp. 403-412.

[5] B. Jessen and A. Wintner, "Distribution functions and the Riemann zeta function," *Transactions of the American Mathematical Society*, vol. 38 (1935), pp. 48-88.

[6] A. Khintchine, "Zu Birkhoffs Lösung des Ergodenproblems," *Mathematische Annalen*, vol. 107 (1933), pp. 485-488.

[7] H. Weyl, "Sur une application de la théorie des nombres à la mécanique statistique et la théorie des perturbations," *Enseignement Mathématique*, vol. 16 (1914), pp. 455-467.

[8] A. Wintner, "Diophantische Approximationen und Hermitesche Matrizen," *Mathematische Zeitschrift*, vol. 30 (1929), pp. 290-319, more particularly pp. 312-316.

[9] A. Wintner, "Sur l'analyse anharmonique des inégalites séculaires fournies par l'approximation de Lagrange," *Rendiconti della R. Accademia Nazionale dei Lincei*, ser. 6, vol. 11 (1930), pp. 464-467.

[10] A. Wintner, "Upon a statistical method in the theory of diophantine approximations," *American Journal of Mathematics*, vol. 55 (1933), pp. 309-331.

ERRATA.

In the paper of E. R. van Kampen and Aurel Wintner, "On a symmetrical canonical reduction of the problem of three bodies," *American Journal of Mathematics*, vol. 59 (1937), pp. 153-166, read 18 instead of 9 in formula (55_2) on p. 165 and in formula (59) on p. 166.

In the paper of E. R. van Kampen and Aurel Wintner, "Convolutions of distributions on convex curves and the Riemann zeta function," *American Journal of Mathematics*, vol. 59 (1937), pp. 175-204,

page	line	instead of	read
188	14	$\epsilon = \bar{\bar{\epsilon}}(\epsilon)$	$\bar{\bar{\epsilon}} = \bar{\epsilon}(\epsilon)$
192	15	tangent	normal
192	16	the	the normal
192	18	$p_m{}^{-2\sigma})^2/$	$p_m{}^{-\sigma})^2/$
193	1	h	k
193	14	$<$	$>$
194	12	Max(	Max($\bar{\sigma}$,

ON AN ABSOLUTE CONSTANT IN THE THEORY OF VARIATIONAL STABILITY.[1]

By E. R. van Kampen and Aurel Wintner.

If $p(t)$, $-\infty < t < +\infty$, is a real continuous periodic function, the linear differential equation

$$\frac{d^2x}{dt^2} + p(t)x = 0 \tag{1}$$

is known to possess two solutions $x = x_1$, $x = x_2$ of the form

$$x_1 = e^{\lambda t/T} f_1(t), \qquad x_2 = e^{-\lambda t/T} f_2(t), \tag{2}$$

where $f_1(t)$ and $f_2(t)$ do not vanish identically and are periodic with the same period T as $p(t)$. The numbers λ and $-\lambda$, the characteristic exponents, are determined mod $2\pi i$ by the reciprocal quadratic equation

$$e^{\pm 2\lambda} - 2Ae^{\pm\lambda} + 1 = 0, \tag{3}$$

this equation being the characteristic equation of a real binary linear substitution of determinant 1. If λ is not a multiple of $T\pi i$, it is clear that the two solutions (2) are linearly independent. If λ is a multiple of $T\pi i$, it depends on the elementary divisors of that binary substitution whether or not the general solution of (1) is free of secular terms. The equation (1) determined by the given periodic function $p(t)$ is said to be of the stable type if every solution $x(t)$ remains bounded as $t \to \pm\infty$, i. e., if the elementary divisors are simple and λ lies on the imaginary axis of the λ-plane. Since from (3)

$$e^{\pm\lambda} = A \pm (A^2 - 1)^{\frac{1}{2}},$$

it follows that

$$-1 \leqq A \leqq 1 \tag{4}$$

is necessary and that

$$-1 < A < 1 \tag{5}$$

is sufficient for the stability of (1). In fact, a multiple elementary divisor is impossible unless there is a double root, so that $A^2 = 1$ is a necessary condition for solutions with secular terms.

The determination of the solutions (2) in terms of $p(t)$ requires, in

[1] Received February 22, 1937.

general, an application of infinite determinants or of equivalent transcendental processes. Even the determination of the characteristic exponents $\pm\lambda$ is quite involved, it being defined by the zeros of Hill's fundamental determinant or by the characteristic equation (3) of the monodromy matrix, i. e., by the number A. The latter can be represented, according to Liapounoff,[2] by means of the convergent series

$$(6) \qquad A = 1 - A_1 + A_2 - A_3 + \cdots,$$

where, if $T > 0$ denotes the period of $p(t)$, the number A_n is the definite integral

$$(7) \qquad A_n = \tfrac{1}{2}\int_0^T dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{n-1}} Q_n(t_1, \cdots, t_n)\, dt_n,$$

while the function Q_n is defined in terms of the primitive function

$$(8) \qquad P(t) = \int_0^t p(\bar{t})\, d\bar{t}$$

of the coefficient $p(t)$ of (1) as follows:

$$(9) \quad Q_n(t_1, \cdots, t_n) \\ = \{P(T) - P(t_1) + P(t_n)\}\{P(t_1) - P(t_2)\} \cdots \{P(t_{n-1}) - P(t_n)\}.$$

In particular, Q_1 is the constant $P(T)$, so that, from (7) and (8),

$$(10) \qquad A_1 = \tfrac{1}{2}TP(T) = \tfrac{1}{2}T \int_0^T p(t)\, dt.$$

Since the actual determination of $\pm\lambda$ requires, in view of (3), (6), (7), (8), (9), highly complicated operations, it is natural to ask for criteria which impose a less remote condition on the coefficient $p(t)$ of (1) and assure that (1) is of the stable type. First, if $p(t)$ is a positive constant, (1) is clearly of the stable type. On the other hand, if the positive periodic function $p(t)$ is not a constant, (1) need not be of the stable type. This holds also when the deviation of $p(t) > 0$ from a constant is less than an arbitrarily small $\epsilon > 0$. Examples to this effect are implied[3] by the theory of Mathieu's equation, where

$$(11) \qquad p(t) = c_1 + c_2 \cos t \qquad (c_1 > c_2 \to 0 < c_1).$$

---

[2] A. Liapounoff, "Sur une série relative à la théorie des équations différentielles linéaires à coefficients périodiques," *Comptes Rendus*, vol. 123 (1896), pp. 1248-1252. This investigation of Liapounoff is reproduced on pp. 425-431 of vol. IV, part III (1902) of Forsyth's *Theory of Differential Equations*.

[3] Cf. M. J. O. Strutt, "Wirbelströme im elliptischen Zylinder," *Annalen der Physik*, vol. 84 (1927), pp. 485-506, where further references are given.

Now there exists, according to Liapounoff,[2] an absolute constant $\alpha > 0$ such that if a real continuous function $p(t)$ of period T is non-negative for every t, positive for some t and satisfies the inequality

$$T\int_0^T p(t)\,dt \leqq \alpha, \tag{12}$$

then (1) is necessarily of the stable type. That α cannot be arbitrarily large, is clear from the example of the Mathieu equation mentioned above. It is indicated by Strutt's diagram,[3] which has been calculated numerically for the boundary curves of the stability region of (11) in the (c_1, c_2)-plane, that the best possible value of the absolute constant α cannot be much greater than 5. On the other hand, every $\alpha \leqq 4$ is admissible. In order to see this, it is, according to Liapounoff, sufficient to observe that, as shown by (8), (9), (7) and (10), the assumption

$$0 \leqq p(t) \not\equiv 0 \tag{13}$$

obviously implies the inequalities

$$A_{n+1} < \frac{A_n A_1}{n+1}, \qquad A_n > 0. \tag{14}$$

In fact, if (12) is satisfied by an $\alpha \leqq 4$, then

$$2 \geqq A_1 > A_2 > A_3 > \cdots > 0$$

in view of (10) and (14), so that the number (6) satisfies the sufficient condition (5) of stability.

Now it will be shown that $\alpha = 4$ is the exact value of the absolute constant in question. Thus there does or does not exist a continuous periodic function $p(t)$ which satisfies (13), (12) and makes (1) a differential equation of the unstable type according as $\alpha > 4$ or $\alpha \leqq 4$. It also will be shown that $\alpha = 4$ remains the greatest admissible value of α also when one restricts $p(t)$ to be an even function of t. Finally, it will be seen from the proof that nothing is gained if one requires that $p(t)$ is analytic or if one replaces (13) by $p(t) > 0$.

Needless to say, the greatest admissible value of α is independent of the value of the period T. This is seen from (12) if one replaces t in (1) by ct, where $c > 0$ is arbitrary. On choosing

$$T = 1, \tag{15}$$

it follows that, since every $\alpha \leqq 4$ is admissible, the statement to be proved may be formulated as follows: There exists for every $\epsilon > 0$ a real non-negative

continuous function $p(t) \not\equiv 0$ of period 1 such that, on the one hand, the mean value of $p(t)$ satisfies the inequality

$$\int_0^1 p(t)\,dt \leqq 4(1+\epsilon) \tag{16}$$

and, on the other hand, the characteristic exponents $\pm\lambda$ of (1) are not of the stable type.

For two fixed numbers β, μ which satisfy the inequalities

$$0 < \mu < \tfrac{1}{2}, \qquad \beta > 0, \tag{17}$$

put

$$\begin{aligned} p(t) &= (\mu - t)\beta/\mu, \text{ if } 0 \leqq t \leqq \mu \text{ and } p(t) = 0, \text{ if } \mu \leqq t \leqq \tfrac{1}{2}, \\ p(t) &= p(1-t), \text{ if } \tfrac{1}{2} < t < 1, \text{ finally } p(t+1) = p(t), \end{aligned} \tag{18}$$

so that $p(t) \not\equiv 0$ is an even, continuous, non-negative function of period 1. It is easy to see that, for this $p(t)$,

$$A_1 = \tfrac{1}{2}\beta\mu \tag{19}$$

and

$$A_2 < \beta^2\mu^3. \tag{20}$$

In fact, (19) is clear from (10), (15) and (18). Furthermore, from (18),

$$p(t) = 0, \text{ if } \mu \leqq t \leqq 1 - \mu;$$

hence, from (8),

$$P(t_1) - P(t_2) = 0, \text{ if } \mu \leqq t_2 \leqq t_1 \leqq 1 - \mu,$$

and so, since

$$Q_2(t_1, t_2) = \{P(1) - P(t_1) + P(t_2)\}\{P(t_1) - P(t_2)\} \tag{21}$$

in view of (9) and (15),

$$Q_2(t_1, t_2) = 0, \text{ if } \mu \leqq t_2 \leqq t_1 \leqq 1 - \mu. \tag{22}$$

Since (8) is a non-decreasing function in view of $p(t) \geqq 0$, it is seen from (21) that

$$Q_2(t_1, t_2) \leqq \{P(1)\}^2, \quad \text{whenever} \quad 0 \leqq t_2 \leqq t_1 \leqq 1.$$

Since $P(1) = \beta\mu$ in view of (10), (15) and (19), it follows that

$$Q_2(t_1, t_2) \leqq \beta^2\mu^2, \quad \text{whenever} \quad 0 \leqq t_2 \leqq t_1 \leqq 1. \tag{23}$$

Now if S denotes that portion of the triangle $0 \leqq t_2 \leqq t_1 \leqq 1$ in the (t_1, t_2)-

plane on which $\mu \leqq t_2 \leqq t_1 \leqq 1 - \mu$ does not hold, then it is seen from (7), (15), (22) and (23) that

$$2A_2 = \int_0^1 \int_0^{t_1} Q_2(t_1, t_2)\, dt_2 dt_1 = \int\int_S Q_2(t_1, t_2)\, dt_2 dt_1 \leqq \beta^2\mu^2 \int\int_S dt_2 dt_1.$$

This proves (20), since, by the definition of the region S,

$$\int\int_S dt_1 dt_2 = \text{area of } S < 2\mu.$$

Now let $\epsilon > 0$ in (16) be given. Since only small values of ϵ need to be considered, one can assume that $\epsilon < \frac{1}{2}$ and then choose the numbers β and μ, which occur in (17) and define the periodic peak function $p(t) \geqq 0$, in such a way that on the one hand

(24) $$\beta\mu = 4(1 + \epsilon)$$

and on the other hand $\mu < \epsilon/(4 + 4\epsilon)^2$. Since the latter inequality implies, in view of (20), that

$$A_2 < \beta^2\mu^2\epsilon/(4 + 4\epsilon)^2,$$

it is seen from (24) and from the assumption $\epsilon < \frac{1}{2}$ that

(25) $$A_2 < \epsilon < \tfrac{1}{2}.$$

On the other hand,

(26) $$A_1 = 2(1 + \epsilon)$$

in view of (19) and (24). Since, by (25), (26) and (14),

$$A_2 > A_3 > A_4 > \cdots > 0,$$

it is clear that

$$0 < A_2 - A_3 + A_4 - \cdots < A_2.$$

Hence $A < 1 - A_1 + A_2$ in view of (6). It follows, therefore, from (25) and (26) that

$$A < 1 - A_1 + \epsilon = 1 - 2(1 + \epsilon) + \epsilon = -1 - \epsilon < -1.$$

Consequently, (4) is not satisfied. Since (4) is a necessary condition for characteristic exponents $\pm\lambda$ of the stable type, and since (16) is satisfied in view of (10), (15) and (26), the proof is complete.

THE JOHNS HOPKINS UNIVERSITY.

ON THE EXPANSION OF THE REMAINDER IN THE OPEN-TYPE NEWTON-COTES QUADRATURE FORMULA.*

By Orville G. Harrold, Jr.

1. The result that the remainder in the Newton-Cotes quadrature formula can be expanded in a series of the Euler-MacLaurin type has been established by J. V. Uspensky.[1] Inasmuch as the open-type quadrature formula of this kind discussed by J. F. Steffensen [2] is important in the numerical integration of differential equations, the question has been raised as to whether or not an analogous development holds for the open formula. The answer is in the affirmative.

2. We consider, without loss of generality, the integration interval $(0, 1)$. The unit interval is divided in n equal parts. The function $f(x)$ to be integrated over this interval is assumed known at $x = 1/n, 2/n, \cdots, (n-1)/n$. The coefficients, or weights, in the quadrature formula will be denoted by A_i:

$$A_i = \int_0^1 \frac{\omega_n(x)\,dx}{\omega'_n(i/n)(x-i/n)},$$
$$\omega_n(x) = (x-1/n)(x-2/n)\cdots(x-(n-1)/n).$$

It is convenient to introduce the symbol

$$K_n{}^\nu = A_1B_\nu(1/n) + A_2B_\nu(2/n) + \cdots + A_{n-1}B_\nu((n-1)/n),$$

where $B_\nu(x)$ is the Bernoullian polynomial of degree ν.

3. With the above notations, the result of this investigation can be formulated more explicitly as follows:

Let $f(x)$ be a continuous function on $0 \leqq x \leqq 1$, with as many continuous derivatives as are needed in the discussion; then

* Received November 10, 1936; revised December 28, 1936.

[1] J. V. Uspensky, "On the expansion of the remainder in the Newton-Cotes formula," *Transactions of the American Mathematical Society*, vol. 37 (1935), pp. 381-396.

[2] J. F. Steffensen, *Interpolation*. Williams and Wilkins Co., Baltimore, 1927.

$$\int_0^1 f(x)\,dx = \sum_{i=1}^{n-1} A_i f(i/n) - \sum_{\nu=m=[n/2]}^{m+s-1} K_n^{2\nu} \frac{\Delta f^{(2\nu-1)}(0)}{(2\nu)\,!} + R_{2s+2m},$$

where

$$R_{2s+2m} = -K_n^{2s+2m} \frac{f^{(2s+2m)}(\xi)}{(2s+2m)\,!}, \qquad 0 < \xi < 1,$$

s being an arbitrary positive integer. Further, if the even ordered derivatives keep their sign on $0 \leqq x \leqq 1$ and if their signs are all alike, then when the series is truncated after a certain term, the error committed is of the same sign as the next term and in absolute value is less than it.

4. By the Euler-Maclaurin summation formula

$$f(x+\theta) = \int_x^{x+1} f(t)\,dt + \sum_{\nu=1}^{r} \frac{B_\nu(\theta)}{\nu\,!} \Delta f^{(\nu-1)}(x) - \int_0^1 \frac{\bar{B}_r(\theta - t)}{r\,!} f^{(r)}(x+t)\,dt, \qquad (0 \leqq \theta \leqq 1).$$

$\bar{B}_\nu(x)$ is the Bernoullian periodic function of order ν, r being an arbitrary positive integer. Fixing $x = 0$ throughout this discussion and allowing θ to take on successively the values $1/n, 2/n, \cdots, (n-1)/n$, we get

$$f(i/n) = \int_0^1 f(t)\,dt + \sum_{\nu=1}^{r} \frac{B_\nu(i/n)}{\nu\,!} \Delta f^{(\nu-1)}(0) - \int_0^1 \frac{\bar{B}_r(i/n - t)}{r\,!} f^{(r)}(t)\,dt, \qquad (i = 1, 2, \cdots, n-1).$$

Multiplying by A_i and summing from $i = 1$ to $n-1$, there results (since $\sum_{i=1}^{n-1} A_i = 1$),

$$\int_0^1 f(t)\,dt = \sum_{i=1}^{n-1} A_i f(i/n) - \sum_{\nu=1}^{r} \frac{K_n^\nu}{\nu\,!} \Delta f^{(\nu-1)}(0) + \int_0^1 \frac{f^{(r)}(t)}{r\,!} \{A_1 \bar{B}_r(1/n - t) + A_2 \bar{B}_r(2/n - t) + \cdots + A_{n-1} \bar{B}_r((n-1)/n - t)\}\,dt.$$

For $\nu \leqq 2m-1$, $m = [n/2]$

$$(*) \qquad K_n^\nu = \int_0^1 B_\nu(x)\,dx = 0.$$

Since $A_i = A_{n-i}$ and $B_{2\nu+1}(i/n) = -B_{2\nu+1}((n-i)/n)$ for all values of ν, we have

$$K_n^{2r+1} = 0.$$

Thus the integral may be presented, for $r = 2s$,

$$\int_0^1 f(t)\,dt = \sum_{k=1}^{n-1} A_k f(k/n) - \sum_{\nu=m}^{s-1} K_n^{2\nu} \frac{\Delta f^{(2\nu-1)}(0)}{(2\nu)!} + R_{2s},$$

where

$$R_{2s} = \int_0^1 \frac{f^{(2s)}(t)}{(2s)!} \{ \sum_{i=1}^{n-1} A_i (\bar{B}_{2s}(i/n - t) - B_{2s}(i/n) \} dt.$$

To establish the result asserted in § 3 it suffices to show:

1°, the numbers $K_n^{2m}, K_n^{2m+2}, \cdots$ alternate in sign;

2°, the quantity $\sum_{i=1}^{n-1} A_i \{\bar{B}_{2s}(i/n - t) - B_{2s}(i/n)\}$ keeps its sign on $0 < t < 1$.

For if 2° is true, we may write, by virtue of (*)

$$R_{2s} = -\frac{f^{(2s)}(\xi)}{(2s)!} \cdot K_n^{2s}, \qquad 0 < \xi < 1.$$

5. To establish the points in question the following properties of

$$Q_k(t) \equiv \sum_{i=1}^{n-1} A_i \{\bar{B}_k(i/n - t) - B_k(i/n)\}$$

are noted:

(1) $Q_k(t) = (-1)^k Q_k(1-t)$;

(2) $Q_k(t)$ is continuous for $k = 2, 3, \cdots$, $Q_k(t)$ possesses derivatives of orders $1, 2, \cdots, k-1$, $Q_k^{(k-2)}(t)$ is not differentiable at $t = i/n$;

(3) $Q''_{2k-1}(t) = (2k-1)(2k-2)Q_{2k-3}(t)$, $\qquad (k = 2, 3, \cdots (t \neq i/n,\ k = 2))$,

$Q'_{2k}(t) = -2kQ_{2k-1}(t) \qquad (k = 1, 2, \cdots (t \neq i/n,\ k = 1))$;

$Q''_{2k}(t) = 2k(2k-1)\{Q_{2k-2}(t) + \sum_{i=1}^{n-1} A_i B_{2k-2}(i/n)\} \qquad (k = 2, 3, \cdots)$;

(4) $Q_k(0) = Q_k(1) = 0, \qquad (k = 1, 2, \cdots)$.

Let α_k and β_k denote respectively the number of distinct zeros of $Q_{2k-1}(t)$ and $Q_{2k}(t)$ on $0 < t < 1$. By virtue of (1), for $k = 1, 2, \cdots$,

(5) $$\alpha_k \geqq 1.$$

From the fact that $Q_{2k}(t)$ has $\beta_k + 2$ distinct zeros on $0 \leqq t \leqq 1$, we get, by the use of (3) and Rolle's theorem, that $Q_{2k-1}(t)$ has at least $\beta_k + 1$ distinct zeros on $0 < t < 1$; thus,

$$\beta_k + 1 \leqq \alpha_k. \tag{6}$$

Due to the fact that $Q_{2k-1}(t)$ has $\alpha_k + 2$ distinct zeros on $0 \leqq t \leqq 1$, we get by (3)' and repeated application of Rolle's theorem that $Q_{2k-3}(t)$ has at least α_k distinct zeros on $0 < t < 1$; hence,

$$\alpha_k \leqq \alpha_{k-1}. \tag{7}$$

If it can be shown that $\alpha_m = 1$, then, by (6) and (7) $\beta_m, \beta_{m+1}, \cdots$ are all zero. Thus, one of our contentions, namely, that

$$\sum_{i=1}^{n-1} A_i\{\bar{B}_{2s}(i/n - t) - B_{2s}(i/n)\}$$

keeps its sign on $0 < t < 1$, will be established.

6. Consider the function $Q_{2m-1}(t) \equiv \sum_{i=1}^{n-1} A_i\{\bar{B}_{2m-1}(i/n - t) - B_{2m-1}(i/n)\}$. Since $\sum_{i=1}^{n-1} A_i B_{2m-1}(i/n) = \int_0^1 B_{2m-1}(x)\,dx = 0$ and $B_m(u+1) - B_m(u) = mu^{m-1}$ we may make the following simplifications:

$$Q_{2m-1}(t) = \sum_{i=1}^{n-1} A_i \bar{B}_{2m-1}(i/n - t) = \sum_{i=1}^{i/n \leqq t} A_i \bar{B}_{2m-1}(i/n - t) + \sum_{i/n > t}^{n-1} A_i \bar{B}_{2m-1}(i/n - t);$$

now

$$\begin{aligned} \sum_{i/n > t}^{n-1} A_i B_{2m-1}(i/n - t) &= \sum_{i/n > t}^{n-1} A_i B_{2m-1}(i/n - t) \\ &= \sum_{i=1}^{n-1} A_i B_{2m-1}(i/n - t) - \sum_{i/n \leqq t} A_i B_{2m-1}(i/n - t), \end{aligned}$$

hence

$$\begin{aligned} Q_{2m-1}(t) &= \sum_{i=1}^{n-1} A_i B_{2m-1}(i/n - t) + \sum_{i/n \leqq t} A_i\{B_{2m-1}(1 + i/n - t) - B_{2m-1}(i/n - t)\} \\ &= \sum_{i=1}^{n-1} A_i B_{2m-1}(i/n - t) + \sum_{i/n \leqq t} A_i (i/n - t)^{2m-2}(2m - 1). \end{aligned}$$

The first term on the right is $\int_0^1 B_{2m-1}(x - t)\,dx = - t^{2m-1}$. By comparison of

$$Q_{2m-1}(t) = - t^{2m-1} + (2m - 1) \sum_{i/n \leqq t} A_i (i/n - t)^{2m-2}$$

with

$$\int_0^1 \phi(t)\,dt = \sum_{i=1}^{n-1} A_i \phi(i/n) + R$$

we see that $R_0(t) \equiv -Q_{2m-1}(t)/(2m-1)$ is precisely the remainder obtained when the open formula is applied to

$$\phi(x) = \begin{cases} (x-t)^{2m-2}, & x \leqq t, \qquad 0 < t < 1, \\ 0, & x > t. \end{cases}$$

Setting

$$R_k(t) = \frac{(-1)^k t^{2m-k-1}}{2m-k-1} - \sum_{i/n \leqq t} A_i (i/n - t)^{2m-k-2},$$

let the number of distinct zeros of $R_k(t)$, for $k = 0, 1, \cdots, 2m-3$ be denoted by $N_0, N_1, \cdots, N_{2m-3}$. Let the number of variations in sign of $R_{2m-2}(t)$, as t varies from 0 to 1 be denoted by N_{2m-2}. Evidently $R_k(t)$ is continuous $(k = 0, 1, 2m-3)$ and possesses a derivative $R'_k(t) = -(2m-k-2)R_{k+1}(t)$, $(k = 0, 1, \cdots, 2m-4)$; but $R_{2m-3}(t)$ does not possess a derivative at $t = i/n$.

By the property of the open quadrature formula

$$R_k(0) = R_k(1) = 0;$$

hence, $R_k(t)$ has $N_k + 2$ distinct zeros on $0 \leqq t \leqq 1$, so that, by Rolle's theorem, and the fact that $R'_k(t) = -(2m-k-2)R_{k+1}(t)$ we get

$$\text{(A)} \qquad N_k + 1 \leqq N_{k+1}$$

from which it follows that

$$N_0 + (2m-3) \leqq N_{2m-3}.$$

In particular, $N_{2m-3} + 1 \leqq N_{2m-2}$. From the fact that the coefficients A_i alternate in sign for $i \leqq m$,[3] it follows that

$$R_{2m-2}(t) = t - \Sigma A_i, \qquad i/n \leqq t,$$

can have at most $2m-1$ variations in sign in $0 < t < 1$ when $n = 2m$. For, we note $R_{2m}(t)$ cannot change sign more than twice in each of the subintervals $(2i-1)/n \leqq t < (2i+1)/n$, while it definitely does not change sign in $0 \leqq t < 1/n$, and has at most one change of sign in $(2m-1)/2m \leqq t < 1$. Hence, for even n,

$$N_{2m-2} \leqq 2m-1,.$$

[3] See Lemma I below $(n \neq 5)$.

from which $N_{2m-3} \leqq 2m - 2$, or $N_0 \leqq +1$. But $N_0 = \alpha_m \geqq +1$; thus $\alpha_m = +1$, $\beta_m = \beta_{m+1} = \cdots = 0$.

If $n = 2m + 1$ we use the relations

$$R_{2m-2}(t) = t - \Sigma A_i, \qquad i/n \leqq t,$$
$$R_k(t) = (-1)^{k-1} R_k(1-t),$$

which follow immediately from the definition of $R_k(t)$.

Two cases are distinguished:

1°. $n = 4k + 1$. As before, $R_{2m-2}(t)$ does not change sign in $0 \leqq t < 1/n$, and not more than twice in each $(2i-1)/n \leqq t < (2i+1)/n$, $i = 1, 2, \cdots, k-1$. It changes sign at most once in $(m-1)/n \leqq t < m/n$ since $A_m = A_{2k}$ is negative by Lemma 1 and $R_{2m-2}(m/n)$ is obviously negative. Thus $R_{2m-2}(t)$ has at most $4(k-1) + 2 + 1 = 2m - 1$ alternations of sign on $0 < t < 1$; hence, $N_{2m-2} \leqq 2m - 1$, $N_0 \leqq +1$, and as before, $N_0 = +1$.

2°. $n = 4k + 3$. Again, $R_{2m-2}(t)$ does not change sign in $0 \leqq t < 1/n$, and not more than twice in each $(2i-1)/n \leqq t < (2i+1)/n$. However, three alternations are possible in $(m-1)/n \leqq t \leqq (m+1)/n$ so that $N_{2m-2} \leqq 2m + 1$. It follows that $R_{2m-3}(t)$ has at most $2m$ zeros on $0 < t < 1$. But $R_{2m-3}(t)$ is an even function of t with respect to $t = 1/2$, so if $R_{2m-3}(t) < 0$ at $t = 1/2$, $N_{2m-3} \equiv 0 \pmod 4$; but $N_{2m-3} \leqq 2m = 4k + 2$, so that $N_{2m-3} \leqq 2m - 2$, hence $N_0 = +1$.

If $R_{2m-3}(1/2) \nless 0$, an impossible situation arises. Since

$$-R_{2m-3}(t) = t^2/2 + \sum_{i/n \leqq t} A_i(i/n - t),$$

our assumption implies that $\Sigma A_i(i/n) \leqq 1/8$, $i/n < 1/2$, which is $(n \neq 3)$ false by Lemma 2 below. Thus in all cases $\alpha_m = 1$, $\beta_m = 0$ $(n \neq 3, 5)$.[4]

7. The functions $Q_{2m}(t), Q_{2m+2}(t), \cdots$ do not change sign on $0 \leqq t \leqq 1$. They are periodic functions with continuous derivatives, and such that $Q'_{2\lambda}(0) = 0$, so that $Q_{2\lambda}(t)$ has the same sign as $Q''_{2\lambda}(0)$. From § 5,

$$Q''_{2m+2k}(0) = \sum_{i=1}^{n-1} A_i B_{2m+2k-2}(i/n),$$

and

$$\int_0^1 Q_{2m+2k}(t)\,dt = -\sum_{i=1}^{n-1} A_i B_{2m+2k}(i/n).$$

[4] 9 deals with the expansions for $n = 3, 5$.

It is thus evident that the coefficients $K_n^{2\nu}$ alternate in sign for fixed n, and $\nu = m, m+1, \cdots$. Both assertions of § 4 have now been established.

8. It remains to establish the two lemmas mentioned in § 6.

LEMMA 1. *The coefficients A_i satisfy the following inequalities for $i \leqq n/2$:*

$$A_i > 0, \qquad i \text{ odd},$$
$$A_i < 0, \qquad i \text{ even},$$

while for $i > n/2$ we use the fact that $A_i = A_{n-i}$.

To demonstrate that the A_i alternate in sign it is convenient to modify the notation slightly to indicate the dependence on n. Set

$$A_k = A_{nk} = (1/n) \int_0^n \frac{P_k(x)}{P_k(k)}\, dx, \qquad (k = 1, 2, \cdots, n-1),$$

where

$$P_k(x) = \frac{(x-1)(x-2)\cdots(x-n+1)}{x-k}.$$

Since $P_k(k) = (-1)^{n-k-1}\Gamma(k)\Gamma(n-k)$,

$$A_{nk} = \frac{(-1)^{n-k-1}}{n\Gamma(k)\Gamma(n-k)} \int_0^n P_k(x)\, dx.$$

Setting $J_n = \int_0^n P_k(x)\, dx$, and recalling well-known formulas for Gamma functions, we find

$$J_n = \frac{(-1)^{n-1}\Gamma(n)}{\pi} \int_0^1 \xi^{-1}(1-\xi)^{n-1} d\xi \int_0^n e^{x \log \xi/(1-\xi)} \frac{\sin \pi x}{x-k}\, dx.$$

The last integral may be split up and presented as

$$\int_0^n e^{x \log \xi/(1-\xi)} \frac{\sin \pi x}{x-k}\, dx = (-1)^k (\xi/(1-\xi))^k \left\{ \int_0^{n-k} e^{t \log \xi/(1-\xi)} \frac{\sin \pi t}{t}\, dt \right.$$
$$\left. + \int_0^k e^{t \log (1-\xi)/\xi} \frac{\sin \pi t}{t}\, dt \right\}.$$

Using the formula

$$\int_0^h e^{ax} \frac{\sin \pi x}{x}\, dx = \pi/2 + \operatorname{arc\,tan} \alpha/\pi + \pi(-1)^{h-1} \int_{-\infty}^{a} \frac{e^{hx}}{\pi^2 + x^2}\, dx,$$

we get

$$\int_0^n e^{x\log\xi/(1-\xi)}\frac{\sin\pi x dx}{x-k}$$
$$=(-1)^k\pi(\xi/(1-\xi))^k\left\{1+(-1)^{n-k-1}\int_{-\infty}^{\log\xi/(1-\xi)}\frac{e^{(n-k)x}dx}{\pi^2+x^2}\right.$$
$$\left.+(-1)^{k-1}\int_{-\infty}^{-\log\xi/(1-\xi)}\frac{e^{kx}dx}{\pi^2+x^2}\right\}.$$

Carrying out the substitution in J_n, we get

$$J_n=(-1)^{n+k-1}\Gamma(k)\Gamma(n-k)$$
$$+\Gamma(n)\left\{\int_0^1\xi^{k-1}(1-\xi)^{n-k-1}d\xi\int_{-\infty}^{\log\xi/(1-\xi)}\frac{e^{(n-k)x}dx}{\pi^2+x^2}\right.$$
$$\left.+(-1)^n\int_0^1\xi^{k-1}(1-\xi)^{n-k-1}d\xi\int_{-\infty}^{-\log\xi/(1-\xi)}\frac{e^{kx}dx}{\pi^2+x^2}\right\}.$$

Since $A_{nk}=\dfrac{(-1)^{n-k-1}}{n}\;\dfrac{J_n}{\Gamma(k)\Gamma(n-k)}$, we can now write

$$A_{nk}=\frac{1}{n}+\frac{n-1}{n}C_{n-2}^{k-1}(-1)^{k+1}\left\{(-1)^n\int_0^1\xi^{k-1}(1-\xi)^{n-k-1}d\xi\int_{-\infty}^{\log\xi/(1-\xi)}\frac{e^{(n-k)x}dx}{\pi^2+x^2}\right.$$
$$\left.+\int_0^1\xi^{k-1}(1-\xi)^{n-k-1}d\xi\int_{-\infty}^{-\log\xi/(1-\xi)}\frac{e^{kx}dx}{\pi^2+x^2}\right\}.$$

The quantity within the brackets becomes, after change of variable in the inner integrals,

$$\int_0^1\xi^{-1}(1-\xi)^{n-1}d\xi\int_0^1\frac{t^{k-1}+(-1)^nt^{n-k-1}}{\pi^2+\log^2(((1-\xi)/\xi)t)}dt.$$

If k is odd and $\leqq n/2$, we see at once that $A_{nk}>0$ for n even or odd. If k is even,

$$A_{nk}=\frac{1}{n}-\left(\frac{n-1}{n}\right)C_{n-2}^{k-1}\left\{\int_0^1\xi^{-1}(1-\xi)^{n-1}d\xi\int_0^1\frac{t^{k-1}+(-1)^nt^{n-k-1}}{\pi^2+\log^2(((1-\xi)/\xi)t)}dt\right\}$$

and, since the integrand is everywhere non-negative for $k\leqq n/2$,

$$-nA_{nk}>-1+(n-1)C_{n-2}^{k-1}\int_0^{1/2}\xi^{-1}(1-\xi)^{n-1}d\xi\int_{\xi/(1-\xi)}^1\frac{t^{k-1}+(-1)^nt^{n-k-1}}{\pi^2+\log^2(((1-\xi)/\xi)t)}$$

or

$$-nA_{nk}>-1+(n-1)C_{n-2}^{k-1}\int_0^{1/2}\frac{\xi^{-1}(1-\xi)^{n-1}d\xi}{\pi^2+\log^2((1-\xi)/\xi)}\int_{\xi/(1-\xi)}^1\{t^{k-1}+(-1)^nt^{n-k-1}$$

On integration of the inner integral

$$-nA_{nk} > -1 + (n-1)C_{n-2}^{k-1}\left\{\frac{1}{k} + \frac{(-1)^n}{n-k}\right\}\int_0^{1/2}\frac{\xi^{-1}(1-\xi)^{n-1}d\xi}{\pi^2 + \log^2((1-\xi)/\xi)}$$

$$-(n-1)C_{n-2}^{k-1}\left\{\frac{1}{k}\int_0^{1/2}\frac{\xi^{k-1}(1-\xi)^{n-k}d\xi}{\pi^2+\log^2((1-\xi)/\xi)} + \frac{1}{n-k}\int_0^{1/2}\frac{\xi^{n-k-1}(1-\xi)^k d\xi}{\pi^2+\log^2((1-\xi)/\xi)}\right\}.$$

The term

$$(n-1)C_{n-2}^{k-1}\left\{\frac{1}{k}\int_0^{1/2}\frac{\xi^{k-1}(1-\xi)^{n-k}d\xi}{\pi^2+\log^2((1-\xi)/\xi)} + \frac{1}{n-k}\int_0^{1/2}\frac{\xi^{n-k-1}(1-\xi)^k d\xi}{\pi^2+\log^2((1-\xi)/\xi)}\right\}$$

is less than

$$\frac{(n-1)}{\pi^2}C_{n-2}^{k-1}\left\{\frac{1}{k}\int_0^{1/2}\xi^{k-1}(1-\xi)^{n-k-1}d\xi + \frac{1}{n-k}\int_0^{1/2}\xi^{n-k-1}(1-\xi)^{k-1}d\xi\right\}$$

which is less than

$$\frac{(n-1)}{\pi^2}C_{n-2}^{k-1}\frac{1}{k}\left\{\int_0^{1/2}\xi^{k-1}(1-\xi)^{n-k-1}d\xi + \int_0^{1/2}\xi^{n-k-1}(1-\xi)^{k-1}d\xi\right\}$$

for $k \leqq n/2$. Hence the term

$$(n-1)\ C_{n-2}^{k-1}\left\{\frac{1}{k}\int_0^{1/2}\frac{\xi^{k-1}(1-\xi)^{n-k}d\xi}{\pi^2+\log^2((1-\xi)/\xi)} + \frac{1}{n-k}\int_0^{1/2}\frac{\xi^{n-k-1}(1-\xi)^k d\xi}{\pi^2+\log^2((1-\xi)/\xi)}\right\}$$

is less than $1/k\pi^2$. Since

$$\int_0^{1/2}\frac{\xi^{-1}(1-\xi)^{n-1}d\xi}{\pi^2+\log^2((1-\xi)/\xi)} \geqq \left(\frac{n}{1+n}\right)^n\int_0^{1/(1+n)}\frac{\xi^{-1}(1-\xi)^{-1}d\xi}{\pi^2+\log^2((1-\xi)/\xi)},$$

we have

$$\int_0^{1/2}\frac{\xi^{-1}(1-\xi)^{n-1}d\xi}{\pi^2+\log^2((1-\xi)/\xi)} \geqq \left(\frac{n}{1+n}\right)^n\frac{1}{\pi}\operatorname{arc\,tan}\left(\frac{\pi}{\log n}\right);$$

hence

$$-nA_{nk} > (n-1)C_{n-2}^{k-1}\left\{\frac{1}{k}+\frac{(-1)^n}{n-k}\right\}\left(\frac{n}{1+n}\right)^n\frac{1}{\pi}\operatorname{arc\,tan}\left(\frac{\pi}{\log n}\right)$$

$$-1-\frac{1}{k\pi^2}\cdots. \tag{B}$$

By calculation, $A_{12,2}$ and $A_{13,2}$ are both negative quantities. Considering n odd and even separately, we see readily that the right side of (B) is positive for $n \geqq 12$, and for even $k \leqq n/2$. Thus for all $k \leqq n/2$, $(n \neq 5)$,

$$A_{n,2l+1} > 0, \qquad A_{n,2l} < 0.$$

For n less than 12 we refer to tables.[2]

LEMMA 2. $\sum_{1/n<\frac{1}{2}} A_i(i/n) > 1/8$, $n \equiv 3 \pmod 4$ $(n > 3)$.

To establish this inequality we start from an expression for A_{nk} used in the proof of the previous lemma:

$$A_{nk} = \frac{1}{n} + \frac{n-1}{n} C_{n-2}^{k-1} \int_0^1 \xi^{-1}(1-\xi)^{n-1} d\xi \int_0^1 \frac{(-t)^{k-1} + (-t)^{n-k-1}}{\pi^2 + \log^2(((1-\xi)/\xi)t)} dt.$$

It follows that

$$\text{(A)} \qquad \sum_{k=1}^{m} A_{nk} k/n = \frac{m(m+1)}{2(2m+1)^2} + \frac{2m}{(2m+1)^2} \int_0^1 \xi^{-1}(1-\xi)^{2m} d\xi$$
$$\times \int_0^1 \sum_{k=1}^{m} \frac{k C_{2m-1}^{k-1}\{(-t)^{k-1} + (-t)^{2m-k}\}}{\pi^2 + \log^2(((1-\xi)/\xi)t)} dt.$$

It suffices to show the double integral (second term) exceeds $\dfrac{1}{8(2m+1)^2}$. To this end we start with the identity, valid for $p + q = 1$,

$$p^m + C_m{}^1 p^{m-1} q + \cdots + C_m{}^\mu p^{m-\mu} q^\mu = \frac{\int_0^p x^{m-\mu-1}(1-x)^\mu dx}{\int_0^1 x^{m-\mu-1}(1-x)^\mu dx}.$$

Setting $x = \dfrac{-y}{1-y}$ in the integrand of the numerator on the right side, we obtain for $m = n-2$, $\mu = \dfrac{n-3}{2}$,

$$p^{n-2} + C^1{}_{n-2} p^{n-3} q + \cdots + C_{n-2}^{(n-3)/2} p^{(n-1)/2} q^{(n-3)/2}$$
$$= (-1)^{(n-1)/2} \frac{\displaystyle\int_0^{-(p/q)} \frac{y^{(n-3)/2} dy}{(1-y)^{n-1}}}{\displaystyle\int_0^1 x^{(n-3)/2}(1-x)^{(n-3)/2} dx};$$

whence, putting $p = -tq$ (so that $q = (1-t)^{-1}$),

$$(-1)^{n-2} q^{n-2} \sum_{k=0}^{(n-3)/2} C^k{}_{n-2} t^{n-k-2} (-1)^k$$
$$= (-1)^{(n-1)/2} \int_0^t \frac{y^{(n-3)/2} dy}{(1-y)^{n-1}} \div \int_0^1 x^{(n-3)/2}(1-x)^{(n-3)/2} dx.$$

Further simplification gives

$$\sum_{k=0}^{m-1} C^k{}_{2m-1}(-1)^k t^{2m-k-1} = mC^m{}_{2m-1}(1-t)^{2m-1}\int_0^t \frac{y^{m-1}dy}{(1-y)^{2m}},$$

which upon differentiation leads to

$$-\sum_{k=0}^{m-1}(k+1)C^k{}_{2m-1}(-1)^k t^{2m-1-k} = -2m\sum_{k=0}^{m-1}C^k{}_{2m-1}(-1)^k t^{2m-1-k}$$
$$+ mC^m{}_{2m-1}\left\{\frac{t^m}{1-t} - t(2m-1)(1-t)^{2m-2}\int_0^t \frac{y^{m-1}dy}{(1-y)^{2m}}\right\}.$$

This relation may be written as

$$-\sum_{k=0}^{m-1}(k+1)C^k{}_{2m-1}(-1)^k t^{2m-k-1}$$
$$= mC^m{}_{2m-1}\left\{\frac{t^m}{1-t} + (t-2m)(1-t)^{2m-2}\int_0^t \frac{y^{m-1}dy}{(1-y)^{2m}}\right\}.$$

To get the other terms, we notice

$$\sum_{k=0}^{2m-1} C^k{}_{2m-1}(-1)^k t^{2m-1-k} = (t-1)^{2m-1},$$

so that

$$\sum_{k=0}^{m-1} C^k{}_{2m-1}(-1)^k t^{k+1} = t(1-t)^{2m-1} + mC^m{}_{2m-1}t(1-t)^{2m-1}\int_0^t \frac{y^{m-1}dy}{(1-y)^{2m}},$$

and, again by differentiation

$$\sum_{k=0}^{m-1}(k+1)(-1)^k C^k{}_{2m-1}t^k = (1-t)^{2m-2}(1-2mt)$$
$$+ mC^m{}_{2m-1}\left\{(1-t)^{2m-2}(1-2mt)\int_0^t \frac{y^{m-1}dy}{(1-y)^{2m}} + \frac{t^m}{1-t}\right\}.$$

Denoting the polynomial in the inner integral in the formula for A_{nk} by $F(t)$, we have

$$\text{(C)}\quad F(t) = \sum_{k=0}^{m-1}(k+1)(-1)^k C^k{}_{2m-1}\{t^k - t^{2m-1-k}\},$$
$$= (1-t)^{2m-2}(1-2mt)$$
$$+ mC^m{}_{2m-1}\left\{\frac{2t^m}{1-t} - (2m-1)(1+t)(1-t)^{2m-2}\int_0^t \frac{y^{m-1}dy}{(1-y)^{2m}}\right\}.$$

Setting

$$\psi(t) = \frac{2t^m}{1-t} - (2m-1)(1+t)(1-t)^{2m-2}\int_0^t \frac{y^{m-1}dy}{(1-y)^{2m}},$$

it is evident that $\psi(0)=0$ and that $\psi(t)>0$ for $0<t<1$, since

$$\left\{\frac{\psi(t)}{(1-t)^{2m-2}(1+t)}\right\}'=\frac{t^{m-1}}{(1+t)^2(1-t)^{2m-2}}>0.$$

Thus, by integration,

$$\begin{aligned}\phi(t)=\int_0^t F(x)\,dx=\int_0^t(1-x)^{2m-2}(1-2mx)\,dx+mC^m{}_{2m-1}\int_0^t\psi(x)\,dx,\\ =t(1-t)^{2m-1}+mC^m{}_{2m-1}\int_0^t\psi(x)\,dx>0,\end{aligned}$$

for $0<t<1$. This expression we use in the integral in (A):

$$\begin{aligned}\int_0^1\frac{F(t)\,dt}{\pi^2+\log^2(((1-\xi)/\xi)t)}&=\frac{\phi(1)}{\pi^2+\log^2((1-\xi)/\xi)}\\ &\quad+\int_0^1\frac{\phi(t)2\log(((1-\xi)/\xi)t)}{[\pi^2+\log^2(((1-\xi)/\xi)t)]^2}\frac{dt}{t},\\ &=\frac{\phi(1)}{\pi^2+\log^2((1-\xi)/\xi)}+R(\xi).\end{aligned}$$

If $1/2<\xi<1$, $\log((1-\xi)/\xi)t<0$; if $0<\xi\leqq 1/2$, $\log((1-\xi)/\xi)t<0$ for $t<\xi/(1-\xi)$. We seek now upper bounds for the absolute values of the negative terms:

$$\begin{aligned}&\int_0^1\frac{\phi(t)2\log(((1-\xi)/\xi)t)}{[\pi^2+\log^2(((1-\xi)/\xi)t)]^2}\frac{dt}{t}, \qquad 1/2<\xi<1;\\ &\int_0^{\xi/(1-\xi)}\frac{\phi(t)2\log(((1-\xi)/\xi)t)}{[\pi^2+\log^2(((1-\xi)/\xi)t)]^2}\frac{dt}{t}, \qquad 0<\xi\leqq 1/2.\end{aligned}$$

Observing that

$$\begin{aligned}\phi(t)=t(1-t)^{2m-1}+mC^m{}_{2m-1}\Bigg\{t^{m-1}\left(\frac{2}{2m-1}-\frac{1}{2m}\right)+t^m\left(\frac{m-1}{2m^2}-\frac{1}{m}\right)\\ +(m-1)(1-t)^{2m-1}\left(\frac{1-t}{2m}-\frac{2}{2m-1}\right)\int_0^t\frac{y^{m-2}\,dy}{(1-y)^{2m-1}}\Bigg\},\end{aligned}$$

we have, since the last two terms of the bracket are negative,

$$\phi(t)<t(1-t)^{2m-1}+\frac{2m+1}{2(2m-1)}C^m{}_{2m-1}t^{m-1}<\frac{1}{2m}+\frac{2m+1}{2(2m-1)}C^m{}_{2m-1},$$

for $0<t<1$. By virtue of these inequalities, the first integral is less in absolute value than

$$\frac{\dfrac{1}{2m}+\dfrac{2m+1}{2(2m-1)}C^m{}_{2m-1}}{\pi^2+\log^2\left(\dfrac{1-\xi}{\xi}\right)},$$

and the second is less than

$$\frac{1}{\pi^2}\left\{\frac{\xi}{1-\xi}+\frac{2m+1}{2(2m-1)}C^m{}_{2m-1}\left(\frac{\xi}{1-\xi}\right)^{m-1}\right\}.$$

Considering

$$\int_0^1 \xi^{-1}(1-\xi)^{2m}\left\{\frac{\phi(1)}{\pi^2+\log^2((1-\xi)/\xi)}+R(\xi)\right\}d\xi,$$

we see that this is greater than

$$\phi(1)\int_0^1\frac{\xi^{-1}(1-\xi)^{2m}d\xi}{\pi^2+\log^2((1-\xi)/\xi)}-\left(\frac{1}{2m}+\frac{2m+1}{2(2m-1)}C^m{}_{2m-1}\right)\int_{1/2}^1\frac{\xi^{-1}(1-\xi)^{2m}d\xi}{\pi^2+\log^2((1-\xi)/\xi)}$$
$$-\frac{1}{\pi^2}\int_0^1\xi^{-1}(1-\xi)^{2m}\left\{\frac{\xi}{1-\xi}+\frac{2m+1}{2(2m-1)}C^m{}_{2m-1}\left(\frac{\xi}{1-\xi}\right)^{m-1}\right\}d\xi,$$

which exceeds

$$\phi(1)\int_0^{1/2}\frac{\xi^{-1}(1-\xi)^{2m}d\xi}{\pi^2+\log^2((1-\xi)/\xi)}$$
$$-\frac{2m+1}{2(2m-1)}C^m{}_{2m-1}\int_{1/2}^1\frac{\xi^{-1}(1-\xi)^{2m}d\xi}{\pi^2+\log^2((1-\xi)/\xi)}-\frac{1}{\pi^2}\int_0^1(1-\xi)^{2m-1}d\xi$$
$$-\frac{(2m+1)}{2\pi^2(2m-1)}C^m{}_{2m-1}\int_0^1\xi^{m-2}(1-\xi)^{m+1}d\xi-\frac{1}{\pi^2m(2m+1)2^{2m+1}}.$$

The second, third, and fourth terms are respectively not greater than

$$\frac{C^m{}_{2m-1}}{\pi^2(2m-1)2^{2m+1}},\quad\frac{1}{2\pi^2m},\quad\text{and}\quad\frac{(m+1)(2m+1)}{4\pi^2m(m-1)(2m-1)}.$$

Since

$$\int_0^{1/2}\frac{\xi^{-1}(1-\xi)^{2m}d\xi}{\pi^2+\log^2((1-\xi)/\xi)}\geqq\left(\frac{2m+1}{2m+2}\right)^{2m+1}\frac{1}{\pi}\operatorname{arc\,tan}\frac{\pi}{\log(2m+1)},$$

we have

$$\int_0^1\xi^{-1}(1-\xi)^{2m}d\xi\int_0^1\frac{F(t)\,dt}{\pi^2+\log^2(((1-\xi)/\xi)t)}$$
$$\geqq\frac{C^m{}_{2m-1}\left(\frac{2m+1}{2m+2}\right)^{2m+1}}{2m(2m-1)}\;\frac{\operatorname{arc\,tan}\frac{\pi}{\log(2m+1)}}{\pi}$$
$$-\frac{C^m{}_{2m-1}}{\pi^2(2m-1)2^{2m+1}}-\frac{1}{\pi^2m}-\frac{(m+1)(2m+1)}{4\pi^2m(m-1)(2m-1)}.$$

The coefficient of the integral is $2m/(2m+1)^2$; hence, we are concerned with the truth of the inequality

$$\frac{2m}{(2m+1)^2}\left\{C^m{}_{2m-1}\left(\frac{\left(\frac{2m+1}{2m+2}\right)^{2m+1}}{2m(2m-1)}\frac{\arctan\left(\frac{\pi}{\log(2m+1)}\right)}{\pi}-\frac{1}{\pi^2(2m-1)2^{2m+1}}\right)\right.$$
$$\left.-\frac{1}{\pi^2 m}-\frac{(m+1)(2m+1)}{4\pi^2 m(m-1)(2m-1)}\right\}>\frac{1}{8(2m+1)^2},$$

or

$$C^m{}_{2m-1}\left\{\frac{\left(\frac{2m+1}{2m+2}\right)^{2m+1}}{2m(2m-1)}\frac{\arctan\frac{\pi}{\log(2m+1)}}{\pi}-\frac{1}{\pi^2(2m-1)2^{2m+1}}\right\}$$
$$>\frac{(m+1)(2m+1)}{4\pi^2 m(m-1)(2m-1)}+\frac{1}{\pi^2 m}+\frac{1}{16m},$$

which is fulfilled for $m \geqq 5$.

The inequality $\Sigma A_i(i/n) > 1/8$ is also fulfilled for $n = 7$ by direct calculation, but not for $n = 3$.

9. The coefficients A_{5k} do not alternate in sign. Noting in particular that (A), § 6, is valid,

$$N_0 + 1 \leqq N_1.$$

Since $N_1 + 1 \leqq N_2$, it follows that $N_0 + 2 \leqq N_2$, where N_2 is the number of sign changes of $R_{2m-2}(t)$ as t increases from 0 to 1 for $m = 2$. From tables the graph of $R_{2m-2}(t) = t - \sum_{i/5 \leq t} A_i$ presents precisely three changes of sign on $0 < t < 1$. Hence,

$$N_0 + 2 \leqq 3,$$

or $N_0 = +1$, which was to be shown.

If $n = 3$, there are only two coefficients, which are equal, and hence there is no question of alternation of sign. The expansion has, in this case, the following form:

$$\int_0^1 f(x)\,dx = \sum_1^2 A_i f(i/3) - \sum_{\nu=1}^{s-1} K_3{}^{2\nu}\frac{\Delta f^{(2\nu-1)}(0)}{(2\nu)!}$$
$$+\int_0^1 \frac{f^{(2s)}(t)}{(2s)!}\sum_1^2 A_i\{\bar{B}_{2s}(i/3-t)-B_{2s}(i/3)\}dt,$$

where

$$K_3{}^{2\nu} = A_1 B_{2\nu}(1/3) + A_2 B_{2\nu}(2/3) = -\frac{(1-3^{1-2\nu})}{2}B_{2\nu},$$

hence, $K_3{}^{2\nu}$ alternates in sign for $\nu = 1, 2, 3, \cdots$.

The mean value theorem can be applied as before to put the remainder in the previously given form since

$$Q_{2s}(t) = \tfrac{1}{2}\{\bar{B}_{2s}(1/3 - t) + \bar{B}_{2s}(2/3 - t) - 2B_{2s}(1/3)\},\ (s = 1, 2, \cdots);$$

does not change sign on $0 < t < 1$. By direct methods this is evident, for on $0 < t < 1/3$,

$$Q_2(t) = t^2,$$

and

$$Q_3(t) = -t^3 + t/6.$$

On $1/3 < t < 2/3$,

$$Q_2(t) = t^2 - t + 1/3,$$
$$Q_3(t) = -t^3 + t/6 + (3/2)(1/3 - t)^2,$$

and thus $\beta_1 = 0$, $\alpha_2 = 1$. From $\beta_s + 1 \leqq \alpha_s$, and $\alpha_s \leqq \alpha_{s-1}$, it is evident that $\beta_2 = \beta_3 = \cdots = 0$.

STANFORD UNIVERSITY,
CALIFORNIA.

ON CERTAIN FUNDAMENTAL IDENTITIES DUE TO USPENSKY.*

By W. A. Dwyer.

1. Introduction. In a series of memoirs entitled "Sur les Relations entre les Nombres des Classes des Formes Quadratiques Binaires et Positives,"[1] Uspensky has obtained several very general fundamental formulae involving incomplete numerical functions in three variables. He has made use of these to establish a great variety of interesting and useful arithmetical theorems.[2] Uspensky's proofs of the fundamental formulae are purely arithmetic and of great simplicity, but give no clue as to how a systematic determination of such formulae may be made. They are in the nature of a priori verifications. The structure of the identities suggests that they may be gotten from equivalent identities involving the theta functions by means of the method of paraphrase.[3] If such an identity can be found, it will suggest a systematic determination, by analytical means, of all identities of this type. From this set of identities it would then be possible to pass back, by means of the method of paraphrase, to other general identities, and then to their application to arithmetic. Bell,[4] for example, has discovered a theta function identity which paraphrases into a certain fundamental formula of Uspensky involving complete numerical functions. In this paper we shall establish two of the formulae involving *incomplete* numerical functions as special cases of a general formula which, in turn, results from the paraphrase of a rather peculiar theta identity.

2. Let $F(x, y, z)$ be a function defined for integral values of the arguments and subject to the parity conditions

$$F(-x, -y, -z) = -F(x, y, z), \qquad F(0, 0, 0) = 0.$$

Then there exists an identity involving *incomplete* numerical functions in three variables

* Received January 26, 1937.

[1] *Bulletin de l'Académie des Sciences de l'U. S. S. R.*, 1925, 1926.

[2] J. V. Uspensky, *loc. cit.*, *Quatrième Memoire*, 1926, pp. 547-566. Also, *American Journal of Mathematics*, vol. 50 (1928), pp. 93-122; *Bulletin of American Mathematical Society*, vol. 36 (1930), pp. 743-754.

[3] E. T. Bell, *Transactions of the American Mathematical Society*, vol. 22 (1921), pp. 1-30, and 198-219.

[4] E. T. Bell, *Bulletin of the American Mathematical Society*, vol. 32 (1926), pp. 682-688.

$$\text{I)}\quad \Sigma F(\delta+i,\delta-d+i,i)$$
$$=\Sigma\left\{F\left(\frac{\Delta'+\Delta}{2},\frac{\Delta'-\Delta}{2},\Delta-h\right)-F\left(-h,\frac{\Delta'-\Delta}{2},\Delta-h\right)\right\}$$
$$+e(n)T+a(n)L,$$
$$T=\sum_{j=1}^{s-1}\{F(j,j,s)-F(s,j,s)\},\qquad L=\Sigma F(r,0,r-t),$$

with integral partitions

II) a) $n=i^2+2d\delta,\quad i\gtreqless 0,\quad \delta>0,\quad d>0,$

b) $n=h^2+\Delta\Delta',\quad h\gtreqless 0,\quad 0<\Delta<\Delta',\quad \Delta'\equiv\Delta\pmod 2,$

c) $n=s^2,\quad s>0,\quad e(n)=0$ or 1 according as n is not or is a perfect square,

d) $n=r^2+t^2,\quad r>0,\quad t>0,\quad a(n)=0$ or 1 according as n is not or is a sum of two squares.

If parity conditions be restricted, corresponding identities result as follows:

$$\text{III)}\quad\begin{cases} F(-x,y,z)=F(x,y,z),\ F(x,-y,-z) \\ \qquad =-F(x,y,z);\ F(x,0,z)=0, \\ \Sigma F(\delta+i,\delta-d+i,i) \\ =\Sigma\left\{F\left(\frac{\Delta'+\Delta}{2},\frac{\Delta'-\Delta}{2},\Delta-h\right)-F\left(h,\frac{\Delta'-\Delta}{2},\Delta-h\right)\right\}+e(n)T,\end{cases}$$

$$\text{IV)}\quad\begin{cases} F(-x,y,z)=-F(x,y,z),F(x,-y,-z)=F(x,y,z),F(x,0,z)=0, \\ \Sigma F(\delta+i,\delta-d+i,i) \\ =\Sigma\left\{F\left(\frac{\Delta'+\Delta}{2},\frac{\Delta'-\Delta}{2},\Delta-h\right)+F\left(h,\frac{\Delta'-\Delta}{2},\Delta-h\right)\right\}+e(n)T,\end{cases}$$

Formulae III and IV are the same as certain formulae discovered by Uspensky [5] and proved by purely arithmetic methods.

3. The theta identity and its paraphrase. Formula I results from the attempt to find a theta-function identity which would paraphrase into III and IV. The procedure consisted in going backwards from these two, and involved the selection of terms, which when arithmetized would meet the conditions II(a) and II(b), and proper adjustment of the arguments. The left side of III or IV suggests the product of a theta and a function of the type

$$\phi_{abc}(x,y)=\frac{\vartheta'_1\vartheta_a(x+y)}{\vartheta_b(x)\vartheta_c(y)}.$$

[5] J. V. Uspensky, *Quatrième Memoire*, *loc. cit.*, and "On incomplete numerical functions," *Bulletin of the American Mathematical Society*, *loc. cit.*, p. 746.

The function desired, and its arithmetic equivalent [6] is

$$\text{V)}\quad \vartheta_3(x+y+z)\phi_{111}(x+y,-y)$$
$$=4\Sigma q^{i^2+2d\delta}\,\Sigma\sin 2[(\delta+i)x+(\delta+i-d)y+iz]$$
$$+\{\operatorname{ctn}(x+y)-\operatorname{ctn}(y)\}\cdot\Sigma\, q^{i^2}\cos 2i(x+y+z).$$

For the terms of III (or IV) involving incomplete numerical functions we shall employ an expression

$$\text{VI)}\quad \chi(x,y)=\sum_{r=-\infty}^{\infty} q^{r^2}e^{-2iry}\operatorname{ctn}(x-r\pi\tau)=\operatorname{ctn}x+2\sum_{n=1}^{\infty}q^{n^2}\sin 2ny$$
$$+4\,\Sigma\, q^n\,\Sigma\sin[(\delta-d)x+2dy],$$
$$0<n=d\delta,\qquad 0<d<\delta,\qquad \delta\equiv d\ (\operatorname{mod}2),$$

which appears as a term in the Fourier development of certain pseudo-periodic functions.[7] By synthesis, we arrived at the relation

$$\text{VII)}\quad \vartheta_3(x+y+z)\phi_{111}(x+y,-y)$$
$$=\vartheta_3(z)\chi(x+y,x+z)-\vartheta_3(x+z)\chi(y,z).$$

An independent proof of this result will appear in § 4 and we shall proceed with the paraphrase of VII. Applying to it the arithmetized expansions of V and VI and the formula $\vartheta_3(x)=\sum_{n=-\infty}^{\infty}q^{n^2}\cos(2nx)$, we obtain

$$\text{VII)}\quad 4\,\Sigma\, q^{i^2+2d\delta}\,\Sigma\sin 2[(\delta+i)x+(\delta-d+i)y+iz]$$
$$=4\,\Sigma\, q^{h^2+\Delta\Delta'}\{\Sigma\sin[(\Delta'+\Delta)x+(\Delta'-\Delta)y+2(\Delta-h)]$$
$$-\Sigma\sin[-2hx+(\Delta'-\Delta)y+2(\Delta-h)z]\}$$
$$+\operatorname{ctn}(x+y)\{\Sigma\, q^{i^2}\{\cos(2iz)-\cos 2i(x+y+z)\}\}$$
$$-\operatorname{ctn}(y)\{\Sigma\, q^{i^2}\{\cos 2i(x+z)-\cos 2i(x+y+z)\}\}$$
$$+2\,\Sigma\, q^{i^2+t^2}\,\Sigma\{\cos(2iz)\sin 2t(x+z)-\cos 2i(x+z)\sin 2tz\},$$

---

[6] Cf. E. T. Bell, "Theta expansions useful in arithmetic," *Messenger of Mathematics*, vol. 53 (1924), pp. 166-176.

[7] M. A. Basoco, "Fourier developments for certain pseudo-periodic functions in two variables," *American Journal of Mathematics*, vol. 54, no. 2 (1932), p. 242. In this connection we shall exhibit the periodicity properties of $\chi(x,y)$. The function is obviously periodic when x or y is increased by π. Furthermore

$$\chi(x+n\pi\tau,y+n\pi\tau)=\chi(x,y),$$
$$\chi(x,y+n\pi\tau)=e^{-2nix}\chi(x,y)$$
$$+ie^{-2nix}\Big\{q^{-n^2}e^{-2ni(y-x)}+2\sum_{k=1}^{n-1}q^{-(n-k)^2}e^{-2i(n-k)(y-x)}+1\Big\}\vartheta_3(y),$$
$$\chi(x+n\pi\tau,y)=q^{n^2}e^{2ni(x-y)}\chi(x,y)$$
$$-i\Big\{q^{n^2}e^{2ni(x-y)}+2\sum_{k=1}^{n-1}q^{n^2-k^2}e^{2i(n-k)(x-y)}+1\Big\}\vartheta_3(y).$$

where the n appearing in the $\sum_{n=1}^{\infty}$ term of VI has been replaced by t, and $i, d, \delta, h, \Delta, \Delta', t$ are subject to the conditions II.

In the terms involving cotangents we change the index of summation from i to s (bringing in the multiplier 2, since $i \gtreqless 0$ while $s > 0$), combine the differences of cosines into the product of two sines, and apply the formula, $\sin(au)\operatorname{ctn}(u) = \sum_{k=0}^{a-1} \cos(a-2k)u$. After combining terms and making an obvious change in the index of summation, our expression contributes

$$+ 4e(n)T + 2\sum q^{t^2}\{\sin 2tz - \sin 2t(x+z)\}.$$

If we split the last term of VIII into a sum corresponding to $i = 0$ and a sum where $i = r$ (r restricted to positive values), we obtain

$$+ 4a(n)L - 2\sum q^{t^2}\{\sin(2tz) - \sin 2t(x+z)\}.$$

Putting the last two results in VIII, changing all arguments to their half-values, and paraphrasing, we arrive at I.

4. Proof of the theta identity. Consider the left-hand side as a function of y alone. Then

$$f(y) = \vartheta_3(x+y+z)\phi_{111}(x+y, -y) = -\frac{\vartheta'_1\vartheta_3(x+y+z)\vartheta_1(x)}{\vartheta_1(x+y)\vartheta_1(y)},$$

$$f(y+n\pi\tau) = q^{n^2} e^{2ni(y-z)} f(y), \qquad f(y+n\pi) = f(y).$$

The residues at the simple poles, $y = 0 + n\pi\tau$, $y = -x + n\pi\tau$, are respectively $-q^{n^2} e^{-2niz}\vartheta_3(x+z)$ and $+q^{n^2} e^{-2ni(x+z)}\vartheta_3(z)$. Let C represent the contour in the y-complex plane composed of $(n+1)$ cells (of width π) above and n cells below the real axis and consider the auxiliary function $\phi(t) = f(t)/\tan(t-y)$ which has poles at $t = y$, $t = y + \pi$, $t = n\pi\tau$, $t = -x + n\pi\tau$. The residue at $t = y$ is $f(y)$. Derange the mesh so that poles lie within the boundary, apply Cauchy's Theorem to $\phi(t)$ around C, and allow *n to become infinite.*[8] Thus,

$$\begin{aligned}\frac{1}{2\pi i}\int_C \phi(t)\,dt = 0 = \sum \text{Residues} \\ &= f(y) + \vartheta_3(x+z)\sum_{n=-\infty}^{\infty} q^{n^2} e^{-2niz}\operatorname{ctn}(y - n\pi\tau) \\ &\quad - \vartheta_3(z)\sum_{n=-\infty}^{\infty} q^{n^2} e^{-2ni(x+z)}\operatorname{ctn}(x+y-n\pi\tau).\end{aligned}$$

The last relation is the same as our identity VII.

---

[8] Cf. M. A. Basoco, "Fourier developments for certain pseudo-periodic functions in two variables," *loc. cit.*, pp. 244-245.

In his arithmetic proof of III and IV Uspensky splits the left-hand side into three sums S_1, S_2, S_3 according as $(\delta - d + i)$ is > 0, < 0, or $= 0$, and sets up a pair of transformations establishing a one-to-one correspondence between the solutions of $i^2 + 2d\delta$ and those of $h^2 + \Delta\Delta'$ which obey the restrictions II(a) and II(b). $S_1 + S_2$, together with *stated parity conditions*, gives us identities III and IV. S_3 obviously vanishes because of the condition $F(x, 0, z) = 0$. If, however, we change the parity conditions to agree with those of I, the more general condition $F(0, 0, 0) = 0$ demands that we consider the contribution of the term S_3. From II(a)

$$n = i^2 + 2d\delta. \quad \text{If } \delta - d + i = 0, \text{ then } n = d^2 + \delta^2.$$

Consequently

$$\Sigma F(\delta + i, \delta + i - d, i)$$

is of the form

$$S_3 = \Sigma F(d, 0, d - \delta),$$

which is of the same form and has the same partitions as the term $a(n)L$ appearing in I.

5. In conclusion, we may point out that the results obtained by Uspensky from his fundamental formulae are implicitly contained in our theta identity. The theta identity suggests other similar products of the theta and ϕ-functions which, when treated in an analogous manner, will lead to general fundamental formulae of the same type as those of Uspensky.

UNIVERSITY OF NEBRASKA.

AN EXTENSION OF BERNSTEIN'S THEOREM ASSOCIATED WITH GENERAL BOUNDARY VALUE PROBLEMS.*

By W. H. McEwen.

Introduction. Consider the n-th order differential system

$$
(1) \qquad \frac{d^n u}{dx^n} + P_2(x) \frac{d^{n-2} u}{dx^{n-2}} + \cdots + P_n(x) u + \lambda u = 0,
$$
$$
U_j(u) = 0, \qquad (j = 1, 2, \cdots, n),
$$

in which the functions $P_2(x), \cdots, P_n(x)$ are continuous and have continuous derivatives of all orders on $a \leqq x \leqq b$, and the U's are n linearly independent conditions involving $u^{(r)}(a), u^{(r)}(b)$, $r = 0, 1, \cdots, n-1$. The general nature of the solutions and the expansion problems connected with this system have been discussed by Birkhoff,[1] Tamarkin,[2] Milne[3] and Stone.[4] Let the characteristic values of the system, taken in the order of magnitude of their moduli, be $\lambda_1, \lambda_2, \cdots$, and let $u_1(x), u_2(x), \cdots$ be the corresponding characteristic solutions. The values λ_k are then the poles of the Green's function of the system. Assume that the boundary conditions are normalized and regular.[5] Assume further that the values λ_k give rise to simple poles of the Green's function when k is large.[6]

Let $S_N(x) = \sum_{j=1}^{N} a_j u_j(x)$ be an arbitrary linear combination of the solutions corresponding to the first N characteristic values, and let L be the maximum value of $| S_N(x) |$ on $a \leqq x \leqq b$. It is the purpose of this paper to establish the following two theorems:

THEOREM 1. *On the interval* $a \leqq x \leqq b$

$$
| S'_N(x) | \leqq q N^2 L,
$$

* Received October 12, 1936; revised January 25, 1937.

[1] G. D. Birkhoff, "Boundary value and expansion problems etc.," *Transactions of the American Mathematical Society*, vol. 9 (1908), pp. 373-395.

[2] J. Tamarkin, *Rendiconti del Circolo di Palermo*, vol. 34 (1912), pp. 345-395.

[3] W. E. Milne, *Transactions of the American Mathematical Society*, vol. 19 (1918), pp. 143-156.

[4] M. H. Stone, "A comparison of the Series of Fourier and Birkhoff," *Transactions of the American Mathematical Society*, vol. 28 (1926), pp. 695-761.

[5] For definitions of these terms see Birkhoff, *loc. cit.*, p. 382.

[6] For a discussion of this assumption see footnote 12.

where q is a positive constant independent of N.

THEOREM 2. *On the interval $a+\delta \leqq x \leqq b-\delta$*

$$| S'_N(x) | \leqq QNL,$$

where Q is a positive constant independent of N.

These theorems are analogous respectively to the theorems of Markoff and Bernstein as applied to polynomial sums. In connection with Theorem 2 it should be noted that the limit QNL cannot in general be extended to the end points a and b, as may be shown by an example,[7] although in certain special cases the limit does apply uniformly to the whole interval. Examples of the latter are the systems that give rise to sums of Fourier or Sturm-Liouville type. The Sturm-Liouville case has been treated by Miss E. Carlson,[8] who has also proved Theorems 1 and 2 for the case of a special 3rd order system.[9]

The proofs given here are based on a number of results to be found in Professor Stone's paper.[10] This paper will be referred to hereafter as (S). The writer wishes to acknowledge his indebtedness to Professor Stone for valuable suggestions in connection with the form of presentation of the proofs.

Preliminary discussion. We can assume, without loss in generality, that the interval of x is $0 \leqq x \leqq 1$, and that the maximum value of $| S_N(x) |$ on $(0,1)$ is 1.

Let $G(x, y; \lambda)$ be the Green's function of system (1). The characteristic values of λ are then the poles of G. The facts concerning the nature and distribution of these values are well known. They form two infinite sequences in the complex λ-plane, given asymptotically by the formulas [11]

$$(2) \qquad \begin{aligned} \lambda_k' &= -(2k\pi i)^n(1+E'/k), \\ \lambda_k'' &= -(-2k\pi i)^n(1+E''/k), \end{aligned}$$

---

[7] The system $du/dx+\lambda u=0$, $u(0)+u(1)=0$ gives rise to sums $S_N(x)$ which are sine and cosine sums in the variable $X=\pi x$ which ranges over a part only of a period interval $0 \leqq X \leqq \pi$. It follows then from the well known form of Bernstein's theorem relating to a trigonometric sum on a part of a period interval, that the limit QNL can be assigned only to an interval which is interior to $0 \leqq X \leqq \pi$, and hence to an interval of x which is interior to $(0,1)$.

[8] E. Carlson, *Transactions of the American Mathematical Society*, vol. 26 (1924), pp. 230-240.

[9] E. Carlson, *Transactions of the American Mathematical Society*, vol. 28 (1926), pp. 435-447; pp. 439-447.

[10] M. H. Stone, *loc. cit.*

[11] See Birkhoff, *loc. cit.*, p. 383.

where E', E'' are bounded functions of k. For large values of k, in accordance with the assumption made on page 1,[12] the poles of G are simple. Hence if multiple poles exist they are limited in number. Let C_N be a circle of the λ-plane with centre at the origin which includes within its boundary the first N poles of G, $\lambda_1, \lambda_2, \cdots, \lambda_N$ and no others. Then the sum $S_N(x)$ may be represented identically by the contour integral[13]

$$(3) \qquad S_N(x) = \frac{1}{2\pi i}\int_0^1 S_N(y)\left[\int_{C_N} G(x, y; \lambda)\, d\lambda\right] dy,$$

provided the poles in question are all simple. If, however, certain of these poles are multiple, the integral given above will represent a sum $\sigma_N(x)$ obtained from $S_N(x)$ by replacing the terms corresponding to the multiple poles λ_s by terms of the form $\int_0^1 S_N(y)\, R_s(x, y)\, dy$, where $R_s(x, y)$ is the residue of G at $\lambda = \lambda_s$. But the terms involved in this change are bounded independently of N, and their number also is independent of N. Hence it follows that S'_N and σ'_N are of the same order of magnitude with respect to N. Thus, for the purpose of our discussion, there is no loss in generality in assuming that $S_N(x)$ is represented by (3).

A more useful form of (3) is obtained by placing $\lambda = \rho^n$. Under this transformation the entire λ-plane is made to correspond to a sector Σ in the ρ-plane, composed of two adjacent sectors of the following set of $2n$ equal sectors:

$$S:\ l\pi/n \leqq \arg\rho \leqq (l+1)\pi/n, \qquad (l = 1, 2, \cdots, 2n-1).$$

The path of integration will then become the arc Γ which the sector Σ cuts off from the circle with centre at the origin in the ρ-plane and radius equal to the n-th root of the radius of C_N. Hence we can write

$$(4) \qquad S_N(x) = \frac{1}{2\pi i}\int_0^1 S_N(y)\left[\int_\Gamma n\rho^{n-1} G(x, y; \rho^n)\, d\rho\right] dy.$$

[12] The assumption referred to is the one which demands that the poles of G be simple when k is large. This condition is not highly restrictive inasmuch as it is automatically satisfied if the system is regular and of odd order, and is in general satisfied if the system is regular and of even order. In the case of even order, however, it may happen that pairs of characteristic values coincide to give double values, and these in turn will give rise to either simple or double poles of G. If the system is self-adjoint the double values give rise to simple poles, but otherwise it is possible to have infinitely many double poles. Tamarkin has given an example of a regular system, $n = 2$, with infinitely many double poles (see Stone and Tamarkin, "Remarks on a paper by Dr. Tautz," *Acta Mathematica*, 1931). It is this type of system which our hypothesis rules out.

[13] See Birkhoff, *loc. cit.*, p. 379.

A special case of (1) is the system (discussed in S, pp. 709-711)

$$d^n u/dx^n + \lambda u = 0, \quad u^{(j)}(0) - u^{(j)}(1) = 0, \qquad (j = 0, 1, \cdots, n-1),$$

which gives rise to sums of Fourier type. Let $\bar{G}(x, y; \lambda)$ denote the Green's function in this case. The arc Γ may be drawn so as to avoid the poles of $\bar{G}$ as well as those of G. Then the integral

$$\text{(5)} \qquad \frac{1}{2\pi i}\int_0^1 S_N(y)\left[\int_\Gamma n\rho^{n-1}\bar{G}(x, y; \rho^n)\,d\rho\right]dy$$

defines a sum $T_N(x)$ which is a trigonometric sum of order $\bar{N}/2$ on the period interval $0 \leqq x \leqq 1$, where $|\bar{N} - N| \leqq K$ independent of the radius of Γ. The latter relation means that $O(\bar{N}/2) = O(N)$.

In regard to the arc Γ we shall demand further that it be kept uniformly away from the poles of G and $\bar{G}$ when the radius is large.

We next define a set of constants which play an important rôle in the asymptotic formulas for G, $\bar{G}$ and their derivatives. These are the n n-th roots of -1, denoted by $\omega_1, \omega_2, \cdots, \omega_n$. For values of ρ on any given sector S let the subscripts be so chosen that

$$R(\rho\omega_1) \leqq R(\rho\omega_2) \leqq \cdots \leqq R(\rho\omega_n), \qquad R \equiv \text{“the real part of.”}$$

Then, if the system is of odd order

$$n = 2\mu - 1,$$

$R(\rho\omega_\mu) = 0$ on the bisecting ray of S, so that in one half of S $R(\rho\omega_\mu) < 0$ whereas in the other half $R(\rho\omega_\mu) > 0$. Let these two halves be denoted by S' and S'' respectively. Thus we have

$$\begin{aligned} &R(\rho\omega_1) \leqq \cdots \leqq R(\rho\omega_\mu) \leqq 0 \leqq R(\rho\omega_{\mu+1}) \leqq \cdots \leqq R(\rho\omega_n) \quad \text{on } S', \\ &R(\rho\omega_1) \leqq \cdots \leqq R(\rho\omega_{\mu-1}) \leqq 0 \leqq R(\rho\omega_\mu) \leqq \cdots \leqq R(\rho\omega_n) \quad \text{on } S''. \end{aligned}$$

On the other hand, if the system is of even order

$$n = 2\mu,$$

$\omega_\mu = -\omega_{\mu+1}$ and on one of the bounding rays of S $R(\rho\omega_\mu) = R(\rho\omega_{\mu+1}) = 0$, so that, throughout the whole of S,

$$R(\rho\omega_1) \leqq \cdots \leqq R(\rho\omega_\mu) \leqq 0 \leqq R(\rho\omega_{\mu+1}) \leqq \cdots \leqq R(\rho\omega_n).$$

These results enable us to state the conditions under which the exponential functions $e^{\rho\omega_j(x-y)}$, $j = 1, 2, \cdots, n$, occurring in the asymptotic formulas for G, $\bar{G}$, etc. are bounded in the form

$$(6) \qquad | e^{\rho\omega_j(x-y)} | \leqq 1, \qquad 0 \leqq x, y \leqq 1,$$

for all values of ρ in question. This inequality holds whenever the real part of $\rho\omega_j(x-y)$ is negative or zero, and hence the specific requirements to be met are as follows:

Case 1. $n = 2\mu - 1$.

$$(6') \qquad j \begin{cases} = 1, \cdots, \mu; & x \geqq y; \quad \rho \text{ on } S', \\ = \mu+1, \cdots, n; & x \leqq y; \quad \rho \text{ on } S'', \\ = 1, \cdots, \mu-1; & x \geqq y; \quad \rho \text{ on } S'', \\ = \mu, \cdots, n; & x \leqq y; \quad \rho \text{ on } S''. \end{cases}$$

Case 2. $n = 2\mu$.

$$(6'') \qquad j \begin{cases} = 1, \cdots, \mu; & x \geqq y; \quad \rho \text{ on } S, \\ = \mu+1, \cdots, n; & x \leqq y; \quad \rho \text{ on } S. \end{cases}$$

In the explicit formulas for G, $\bar{G}$, etc., which will be used presently, it will be seen that these conditions are satisfied in every instance, so that the exponentials occurring in these formulas are bounded in the manner of (6).

We now observe a number of lemmas, the essential parts of the proofs of which are based on results found in (S). The notation $\{A; B\}$ is used to indicate that A is to be taken if $x \geqq y$, and B if $x \leqq y$. The letter R is used to denote the radius of the circular arc Γ, so that for values of ρ on Γ $|\rho| = R$. In outlining the proofs it is necessary to treat separately the cases $n = 2\mu - 1$ and $n = 2\mu$, inasmuch as the asymptotic formulas concerned are different in these two cases. It is sufficient, however, to treat only one of the two equal sectors S which make up Σ. The part of Γ belonging to S will be denoted by γ, and the two halves of it corresponding to S', S'' by γ', γ'' respectively.

LEMMA 1. *For values of* ρ *on* Γ

$$n\rho^{n-1}\left\{\frac{\partial G}{\partial x}; \frac{\partial G}{\partial x}\right\} = O(R)$$

uniformly on $0 \leqq x, y \leqq 1$.

Case 1. $n = 2\mu - 1$. For values of ρ on γ' the following asymptotic formula for the expression in question is given in (S, p. 745, with $k = 1$):

$$(7) \quad n\rho^{n-1}\left\{\frac{\partial G}{\partial x}; \frac{\partial G}{\partial x}\right\}$$
$$= \left\{ F_{11}{}^0 - \sum_{j=1}^{\mu} e^{\rho\omega_j(x-y)} \frac{m_j(x,y,\rho)}{\rho}; \; F_{11}{}^1 + \sum_{j=\mu+1}^{n} e^{\rho\omega_j(x-y)} \frac{m_j(x,y,\rho)}{\rho} \right\}$$
$$+ \frac{\Delta_1{}^{(1)}}{[\theta_0] + e^{\rho\omega_\mu}[\theta_1]} \cdot$$

where

$$F_{11}{}^0 \equiv -\sum_{j=1}^{\mu} \omega_j e^{\rho\omega_j(x-y)}(A_{11}(x) + B_1(y) + \rho\omega_j),$$

$$F_{11}{}^1 \equiv +\sum_{j=\mu+1}^{n} \omega_j e^{\rho\omega_j(x-y)}(A_{11}(x) + B_1(y) + \rho\omega_j),$$

$$([\theta_0] + e^{\rho\omega_\mu}[\theta_1])^{-1} = O(1),^{14}$$

and $\Delta_1{}^{(1)}$ is a determinant of order $n+1$, which, if expanded according to the elements of the first row, has the form

$$\Delta_1{}^{(1)} \equiv \sum_{j=1}^{\mu} (\rho\omega_j) e^{\rho\omega_j x} Q_j(x, y, \rho) + \sum_{j=\mu+1}^{n} (\rho\omega_j) e^{\rho\omega_j(x-1)} Q_j(x, y, \rho).$$

The functions $m_j(x, y, \rho)$, $A_{11}(x)$, $B_1(y)$, $Q_j(x, y, \rho)$ are bounded in their respective variables on $0 \leqq x, y \leqq 1$ when R is large. A similar formula holds when ρ is on γ'', the only difference being that the summations are extended over the ranges $(1, \mu-1)$ and (μ, n).

On examining the individual terms in these expressions it is seen that conditions (6′) are satisfied, so that the exponentials are all bounded in the manner of (6). Hence the terms are either $O(1)$ or $O(R)$ for values of ρ on γ' and γ''; that is, $n\rho^{n-1}\left\{\frac{\partial G}{\partial x}; \frac{\partial G}{\partial x}\right\} = O(R)$ on γ.

Case 2. $n = 2\mu$. The asymptotic formula for $n\rho^{n-1}\left\{\frac{\partial G}{\partial x}; \frac{\partial G}{\partial x}\right\}$ when ρ is on γ is similar in form to the one used in Case 1, ρ on S'; an explicit expression for it is given in (S, p. 760, with $k = 1$). The exponentials involved satisfy conditions (6″) and so are bounded as in (6), and the steps of the proof go through exactly as in Case 1.

LEMMA 2. $\int_\Gamma n\rho^{n-1}[G(x, y; \rho^n) - G(x, y; \rho^n)]d\rho = O(1)$ uniformly on $0 < \delta \leqq x \leqq 1 - \delta$, $0 \leqq y \leqq 1$.

Case 1. $n = 2\mu - 1$. This is (S, Theorem VII, p. 716, with $l = 0$).

Case 2. $n = 2\mu$. The proof in this case is analogous to that of the theorem cited above. In brief outline it is as follows:

For ρ on γ we have, by (S, p. 755),

---

[14] The arc Γ must be kept uniformly away from the poles of G when R is large, so as to make this fraction bounded for large values of R. In view of the manner of distribution of the characteristic values it is clear that this can always be done.

$$n\rho^{n-1}G(x, y; \rho^n) \equiv \{-\sum_{j=1}^{\mu} e^{\rho\omega_j(x-y)}[\omega_j]; + \sum_{j=\mu+1}^{n} e^{\rho\omega_j(x-y)}[\omega_j]\}\text{ [15]}$$
$$+ \frac{\Delta_3}{[\theta_1]e^{2\rho\omega_\mu} + [\theta_0]e^{\rho\omega_\mu} + [\theta_2]}.$$

The arc γ being kept uniformly away from the poles of G when R is large, the denominator of the second term is bounded away from zero. The numerator of the second term is a determinant of order $n+1$, which, if expanded according to the elements of the first row, has the form of a linear combination of the functions $e^{\rho\omega_j x}$ $(j = 1, \cdots, \mu)$, $e^{\rho\omega_j(x-1)}$ $(j = \mu+1, \cdots, n)$ with coefficients which are bounded functions of x, y, ρ. The second term above may thus be written in the form

$$\sum_{j=1}^{\mu} e^{\rho\omega_j x} M_j(x, y, \rho) + \sum_{j=\mu+1}^{n} e^{\rho\omega_j(x-1)} M_j(x, y, \rho),$$

in which the functions $M_j(x, y, \rho)$ are uniformly bounded on $0 \leqq x, y \leqq 1$ for large values of R.

In the special case of system (1) which defines $\bar{G}$, a corresponding formula will represent $n\rho^{n-1}\bar{G}$ on γ. It will be identical with the one given above except for different terms of higher order in the asymptotic forms $[\omega_i]$ and different functions $\bar{M}_i(x, y, \rho)$. Hence on subtracting these results we obtain

$$n\rho^{n-1}(G - \bar{G}) \equiv \left\{-\sum_{j=1}^{\mu} e^{\rho\omega_j(x-y)} \frac{a_j}{\rho}; + \sum_{j=\mu+1}^{n} e^{\rho\omega_j(x-y)} \frac{a_j}{\rho}\right\}$$
$$+ \sum_{j=1}^{\mu} e^{\rho\omega_j x}(M_j - \bar{M}_j) + \sum_{j=\mu+1}^{n} e^{\rho\omega_j(x-1)}(M_j - \bar{M}_j).$$

The exponentials in $\{\ \}$ are bounded as in (6). Hence $\int_\gamma n\rho^{n-1}(G - \bar{G})\,d\rho$ is expressible in terms of integrals of the form

$$\int_\gamma \frac{m\,d\rho}{\rho},\quad \int_\gamma e^{\rho\omega_j x} m\,d\rho \quad (j = 1, \cdots, \mu),\quad \int_\gamma e^{\rho\omega_j(x-1)} m\,d\rho \quad (j = \mu+1, \cdots, n),$$

which, according to (S, Lemmas III, IV′, V′, p. 714 and pp. 754-755, with $k = l = 0$) are uniformly bounded on $0 < \delta \leqq x \leqq 1 - \delta$.

LEMMA 3. $\int_\Gamma n\rho^{n-1}\left\{\frac{\partial G}{\partial x} - \frac{\partial \bar{G}}{\partial x}; \frac{\partial G}{\partial x} - \frac{\partial \bar{G}}{\partial x}\right\} d\rho = O(R)$ *uniformly on* $0 < \delta \leqq x \leqq 1 - \delta,\ 0 \leqq y \leqq 1.$

---

[15] The notation $[W]$ indicates an asymptotic form in ρ in which W is the leading term.

Case 1. $n = 2\mu - 1$. For ρ on γ' we use again formula (7) of Lemma 1, writing $E_j(x, y, \rho)$ for $Q_j(x, y, \rho)/([\theta_0] + e^{\rho\omega_\mu}[\theta_1])$.

$$n\rho^{n-1}\left\{\frac{\partial G}{\partial x}; \frac{\partial G}{\partial x}\right\} \equiv \left\{F_{11}{}^0 - \sum_{j=1}^{\mu} e^{\rho\omega_j(x-y)}\frac{m_j}{\rho};\ F_{11}{}^1 + \sum_{j=\mu+1}^{n} e^{\rho\omega_j(x-y)}\frac{m_j}{\rho}\right\} + \sum_{j=1}^{\mu}(\rho\omega_j)e^{\rho\omega_j x}E_j + \sum_{j=\mu+1}^{n}(\rho\omega_j)e^{\rho\omega_j(x-1)}E_j.$$

A similar form holds for $\bar{G}$, with $F_{11}{}^0$, $F_{11}{}^1$, m_j, E_j replaced by $\bar{F}_{11}{}^0$, $\bar{F}_{11}{}^1$, $\bar{m}_j$, $\bar{E}_j$ respectively. Hence on forming the difference of these two formuals we obtain

$$\begin{aligned} n\rho^{n-1}&\left\{\frac{\partial G}{\partial x} - \frac{\partial \bar{G}}{\partial x}; \frac{\partial G}{\partial x} - \frac{\partial \bar{G}}{\partial x}\right\} \\ &\equiv \left\{(F_{11}{}^0 - \bar{F}_{11}{}^0) - \sum_{j=1}^{\mu} e^{\rho\omega_j(x-y)}\frac{(m_j - \bar{m}_j)}{\rho}\right. \\ &\quad \left.;\ (F_{11}{}^1 - \bar{F}_{11}{}^1) + \sum_{j=\mu+1}^{n} e^{\rho\omega_j(x-y)}\frac{(m_j - \bar{m}_j)}{\rho}\right\} \\ &\quad + \sum_{j=1}^{\mu}(\rho\omega_j)e^{\rho\omega_j x}(E_j - \bar{E}_j) + \sum_{j=\mu+1}^{n}(\rho\omega_j)e^{\rho\omega_j(x-1)}(E_j - \bar{E}_j). \end{aligned}$$

But

$$F_{11}{}^0 \equiv -\sum_{j=1}^{\mu}\omega_j e^{\rho\omega_j(x-y)}(A_{11}(x) + B_1(y) + \rho\omega_j),$$

whereas, according to (S, p. 746, with $s = 1$),

$$\bar{F}_{11}{}^0 \equiv -\sum_{j=1}^{\mu}\omega_j e^{\rho\omega_j(x-y)}(\rho\omega_j).$$

Hence, by (6),

$$F_{11}{}^0 - \bar{F}_{11}{}^0 \equiv -\sum_{j=1}^{\mu}\omega_j e^{\rho\omega_j(x-y)}(A_{11}(x) + B_1(y)) = O(1).$$

Likewise $F_{11}{}^1 - \bar{F}_{11}{}^1 = O(1)$. The remaining terms of $\{\ \}$ are similarly bounded (approaching zero as $R \to \infty$). Hence the integral of the expression in $\{\ \}$ taken over the arc γ' will be of the order R. Moreover the integrals

$$\int_{\gamma'}\rho e^{\rho\omega_j x}(E_j - \bar{E}_j)\,d\rho \qquad (j = 1, \cdots, \mu - 1),$$

$$\int_{\gamma'}\rho e^{\rho\omega_j(x-1)}(E_j - \bar{E}_j)\,d\rho \qquad (j = \mu + 1, \cdots, n)$$

converge to zero uniformly on $0 < \delta \leqq x \leqq 1 - \delta$ as $R \to \infty$, in accordance with (S, Lemma IV, p. 714, with $k = 1$). Hence

$$\int_{\gamma'} n\rho^{n-1}\left\{\frac{\partial G}{\partial x} - \frac{\partial \bar{G}}{\partial x}; \frac{\partial G}{\partial x} - \frac{\partial \bar{G}}{\partial x}\right\}d\rho = O(R) + \omega_\mu\int_{\gamma'}\rho e^{\rho\omega_\mu x}(E_\mu - \bar{E}_\mu)\,d\rho.$$

In a like manner we find that the integral of this expression over γ'' is of the form

$$O(R) + \omega_\mu \int_{\gamma''} \rho e^{\rho\omega_\mu(x-1)} (E'_\mu - \bar{E}'_\mu)\, d\rho.$$

It remains now for us to determine the order of the second terms in the two expressions written above. The integrals involved may be put in the form

$$R \int_{\gamma'} e^{\rho\omega_\mu x} m'\, d\rho, \qquad R \int_{\gamma''} e^{\rho\omega_\mu(x-1)} m''\, d\rho,$$

where $m' = \rho(E_\mu - \bar{E}_\mu)/R = O(1)$ on γ', and $m'' = \rho(E'_\mu - \bar{E}'_\mu)/R = O(1)$ on γ''. An application of (S, Lemma V, p. 714, with $k = l = 0$) will then show that the multipliers of R are uniformly bounded on $0 < \delta \leqq x \leqq 1 - \delta$, and hence we conclude that the second terms also are of order R in this interval.

Case 2. $n = 2\mu$. The argument in this case, based on appropriate formulas found in (S), is entirely similar to the one just given.

LEMMA 4. *The number of poles of $G(x, y; \rho^n)$ on the sector Σ enclosed by the arc Γ is given asymptotically by*

$$N \sim \frac{1}{\pi} R,$$

(*each pole being counted according to its multiplicity*).

This is immediately evident from formulas (2) which give the distribution of the characteristic values of $\lambda = \rho^n$. From this lemma it follows that

$$O(R) = O(N).$$

The proofs of Theorems 1 and 2. On differentiating with respect to x in formula (4) we get

$$S'_N(x) = \frac{1}{2\pi i} \int_0^1 S_N(y) \left[\int_\Gamma n\rho^{n-1} \left\{ \frac{\partial G}{\partial x}; \frac{\partial G}{\partial x} \right\} d\rho \right] dy.$$

But $| S_N(x) | \leqq 1$ on $0 \leqq y \leqq 1$. Hence, by Lemmas 1 and 4,

$$S'_N(x) = \frac{1}{2\pi i} \int_0^1 S_N(y) \left[\int_\Gamma O(R)\, d\rho \right] dy = O(R^2) = O(N^2)$$

on $0 \leqq x \leqq 1$. This proves Theorem 1.

Next, let us consider the trigonometric sum $T_N(x)$ defined by (5). On subtracting it from $S_N(x)$ we have, by reason of Lemma 2,

$$S_N(x) - T_N(x) = \frac{1}{2\pi i}\int_0^1 S_N(y)\left[\int_\Gamma n\rho^{n-1}(G-\bar{G})\,d\rho\right]dy$$
$$= \frac{1}{2\pi i}\int_0^1 S_N(y)\,O(1)\,dy = O(1)$$

on $0 < \delta \leqq x \leqq 1-\delta$. But $|S_N(x)| \leqq 1$, and hence

$$T_N(x) = O(1)$$

on the interval $0 < \delta \leqq x \leqq 1-\delta$, which is interior to the period interval $(0,1)$. It follows then, from a special form of Bernstein's theorem given by D. Jackson,[16] that

$$T'_N(x) = O(\bar{N}/2) = O(N) \tag{8}$$

uniformly on this interior interval.

Finally, on differentiating with respect to x in the formula for $S_N(x) - T_N(x)$ we obtain, by the help of Lemmas 3 and 4,

$$S'_N(x) - T'_N(x) = \frac{1}{2\pi i}\int_0^1 S_N(y)\left[\int_\Gamma n\rho^{n-1}\left\{\frac{\partial G}{\partial x} - \frac{\partial \bar{G}}{\partial x};\ \frac{\partial G}{\partial x} - \frac{\partial \bar{G}}{\partial x}\right\}d\rho\right]dy$$
$$= \frac{1}{2\pi i}\int_0^1 S_N(y)\,O(R)\,dy = O(R) = O(N)$$

uniformly on $0 < \delta \leqq x \leqq 1-\delta$. Hence, by (8),

$$S'_N(x) = T'_N(x) + O(N) = O(N)$$

uniformly on $0 < \delta \leqq x \leqq 1-\delta$. This proves Theorem 2.

Application to a problem of best approximation. Let $f(x)$ be a given function continuous on $a \leqq x \leqq b$, and let it be required to define for each positive integral value of N a function of best approximation to $f(x)$ of the form

$$S_N(x) = \sum_{k=1}^{N} a_k u_k(x),$$

in which the u's are the characteristic solutions of system (1). This may be done by adopting as a measure of approximation the integral

$$\int_a^b |f(x) - S_N(x)|^r\,dx,$$

where r is any given real constant > 0, and requiring that the coefficients of

[16] D. Jackson, *Transactions of the American Mathematical Society*, vol. 26 (1924), pp. 133-154; p. 145.

$S_N(x)$ be chosen in such a way as to give a minimum value to this integral. It is well known that such determinations of the coefficients can always be made, and when $r > 1$ the result is unique.

The question of the convergence of $S_N(x)$ to $f(x)$ as N becomes infinite may be investigated by methods similar to those used by Jackson[17] in the study of the corresponding problems relating to trigonometric sums and polynomials. These methods involve in an essential way the use of Theorems 1 and 2, and lead to the following general theorem:

THEOREM 3. *If $\pi_N(x)$ be an arbitrary sum of the u's of order N and h_N be the maximum value of $|f(x) - \pi_N(x)|$ on $a \leqq x \leqq b$, then there will exist positive constants C_1 and C_2 independent of N such that*

(a) *on* $a \leqq x \leqq b$, $|f(x) - S_N(x)| \leqq C_1 N^{2/r} h_N$,

(b) *on* $a + \delta \leqq x \leqq b - \delta$, $|f(x) - S_N(x)| \leqq C_2 N^{1/r} h_N$.

Thus the question of convergence is made to depend directly on the degree of approximation represented by h_N, that is on the degree of approximation to $f(x)$ that is possible by sums of the form $\pi_N(x)$. In this connection we have the theorems on the degree of convergence of Birkoff's series given by Milne[18] which enable us to state explicit hypotheses under which the quantities $N^{2/r}h_N$ and $N^{1/r}h_N$ will converge to zero as N becomes infinite. The following theorem is given as typical of what can be done in this direction:

THEOREM. *In the case $r > 1$, if $f(x)$ has a first derivative of limited variation on $a \leqq x \leqq b$, and if $f(x)$ vanishes at a and b, then $h_N = O(1/N)$ so that $S_N(x)$ converges uniformly to $f(x)$ on the sub-interval $a + \delta \leqq x \leqq b - \delta$ as N becomes infinite.*

MOUNT ALLISON UNIVERSITY,
SACKVILLE, N. B., CANADA.

[17] D. Jackson, *American Mathematical Society, Colloquium Publications*, vol. 11, pp. 92-101; see also D. Jackson, *Bulletin of the American Mathematical Society*, December, 1933, pp. 889-906.

[18] W. E. Milne, *loc. cit.*, pp. 154-156.

ABSTRACT COVARIANT VECTOR FIELDS IN A GENERAL ABSOLUTE CALCULUS.*

By A. D. MICHAL.

Introduction. The elements of a general absolute differential calculus based on a linear connection and the notion of a contravariant vector field only has recently been considered by me.[1] In the present paper additional postulates on the transformation of Banach "coördinates" are considered and a linear connection of *covariant type* is postulated. A brief treatment is then given of a general absolute differential calculus based on covariant vector fields as well as on contravariant vector fields. The ideas centering around the various adjoints of the Fréchet differentials [2] are of fundamental importance here as well as in the instances in which the Banach space is an infinitely dimensional function space.

1. Abstract coördinate transformations. Let E be a Banach space in which there exists a function $[x, y]$ with the following properties: [3]

(1) $[x, y]$ is a bilinear function on E^2 to the real numbers
(2) $[x, y] = [y, x]$
(3) $[x, y]$ is positive definite; i. e., $[x, x] \geqq 0$ and $[x, x] = 0$ if and only if $x = 0$.

DEFINITION. *A function $T^*(\xi)$ on E to E will be said to be the adjoint of a linear function $T(\xi)$ on E to E if*

(1) *$T^*(\xi)$ is a linear function*
(2) $[T(\xi), \eta] = [\xi, T^*(\eta)]$.

Let U_0 be a fixed Hausdorff neighborhood of a Hausdorff [4] space T. We

* Presented to the American Mathematical Society under a different title, April, 1936. Received by the Editors February 16, 1937.

[1] Michal (1).

[2] By the notation $f(x, y_1, \cdots, y_n; \lambda)$ we shall always mean the partial Fréchet differential of $f(x, y_1, \cdots, y_n)$ in x with increment λ. Occasionally we shall write $\delta\phi(x)$ for $\phi(x; \delta x)$ and $\overset{x}{\underset{\delta x}{dy}} f(x, y)$ for the partial Fréchet differential of $f(x, y)$ in y evaluated at $y = x$.

[3] For motions and rotations in such spaces see Michal, Highberg and Taylor (1).

[4] More generally one can take a *Fréchet neighborhood space* V and require the coördinate systems to be merely reciprocal (1-1) transformations.

shall assume that there exists an open set $S \subset E$ that is a homeomorphic map of U_0. We postulate the existence of coördinate systems $x(P)$: homeomorphic correspondences mapping Hausdorff neighborhoods onto open sets $\Sigma \subset S$. Suppose $x(P)$ and $\bar{x}(P)$ are coördinate systems for two intersecting Hausdorff neighborhoods U_1 and U_2 respectively and let Σ_1 and Σ_2 be the respective maps in S. Then the intersection of U_1 and U_2 induces a homeomorphic mapping, called a coördinate transformation, of an open subset S_1 of Σ_1 onto an open subset S_2 of Σ_2. We shall denote this coördinate transformation by $\bar{x} = \bar{x}(x)$.

It is convenient to call the Hausdorff neighborhood and the map $\Sigma \subset S$, the geometrical domain and the coördinate domain respectively of the coördinate system. We shall assume that each coördinate transformation $\bar{x}(x)$ and its inverse $x(\bar{x})$ have Fréchet differentials $\bar{x}(x; \delta x)$ and $x(\bar{x}; \delta\bar{x})$ throughout the sub-coördinate domains S_1 and S_2 respectively of the coördinate systems $x(P)$ and $\bar{x}(P)$. We shall further assume that $\bar{x}(x; \delta x)$ possesses an adjoint $\bar{x}^*(x; \delta x)$ and that $x(\bar{x}; \delta\bar{x})$ has an adjoint $x^*(\bar{x}; \delta\bar{x})$. It can be shown readily that $\bar{x}(x; \delta x)$ is a solvable linear function of δx with $x(\bar{x}; \delta\bar{x})$ as inverse. From the postulates for $[x, y]$ and the following evident steps

$$\begin{aligned}[\delta\bar{x}, \xi] = [\bar{x}(x; x(\bar{x}; \delta\bar{x})), \xi] &= [x(\bar{x}; \delta\bar{x}), \bar{x}^*(x; \xi)] \\ &= [\delta\bar{x}, x^*(\bar{x}; \bar{x}^*(x; \xi))]\end{aligned}$$

it follows that

$$x^*(\bar{x}; \bar{x}^*(x; \xi)) = \xi \quad \text{for all } \xi \epsilon E. \tag{1.1}$$

Similarly

$$\bar{x}^*(x; x^*(\bar{x}; \xi)) = \xi \quad \text{for all } \xi \epsilon E. \tag{1.2}$$

From these two results it follows that $\bar{x}^*(x; \delta x)$ is a solvable linear function of δx with $x^*(\bar{x}; \delta\bar{x})$ as inverse.

2. Covariant differential of a covariant vector field. The absolute calculus of *contravariant* vector fields has been studied elsewhere.[5] The components of a geometric object have a characteristic law of transformation in the intersection of two Hausdorff neighborhoods. The law of transformation for a contravariant vector field is

$$\bar{\xi}(\bar{x}) = \bar{x}(x; \xi(x)). \tag{2.1}$$

DEFINITION 2.1. *A covariant vector field is a geometric object whose components transform in the intersection of two Hausdorff neighborhoods according to the law*

$$\bar{\eta}(\bar{x}) = x^*(\bar{x}; \eta(x)) \tag{2.2}$$

under a transformation of coördinates $\bar{x} = \bar{x}(x)$.

[5] Michal (1).

The inverse of (2.2) is

$$\eta(x) = \bar{x}^*(x; \bar{\eta}(\bar{x})). \tag{2.3}$$

In addition to the restrictions of § 1 we shall now assume that the second Fréchet differential $\bar{x}(x; \delta_1 x; \delta_2 x)$ exists continuous in x and that the Fréchet differential

$$\underset{\delta\bar{x}}{d_\mu^{\bar{x}}}\, x^*(\sigma; \eta)$$

exists continuous in $\bar{x}$. It can be shown with the aid of theorems proved elsewhere[6] that $x(\bar{x}; \delta_1\bar{x}; \delta_2\bar{x})$ exists continuous in $\bar{x}$. It can also be shown that the adjoint $x^*(\bar{x}; \mu; \lambda)$ of the linear function $x(\bar{x}; \mu; \lambda)$ of λ exists continuous in $\bar{x}$ and that the adjoint $x^*(\bar{x}, \lambda, \mu)$ of the linear function $\underset{\mu}{d_\sigma^{\bar{x}}}\, x^*(\sigma; \lambda)$ of μ exists continuous in $\bar{x}$.

THEOREM 2.1. *Under the restrictions on coördinate transformations just described, the following relations are valid*

$$\underset{\mu}{d_\sigma^{\bar{x}}}\, x^*(\sigma; \lambda) = x^*(\bar{x}; \mu; \lambda) \tag{2.4}$$

$$\underset{\mu}{d_\sigma^{\bar{x}}}\, x^*(\sigma; \lambda) = x^*(\bar{x}, \lambda, \mu). \tag{2.5}$$

The functions $x^*(\bar{x}; \mu; \lambda)$ and $x^*(\bar{x}, \lambda, \mu)$ are bilinear in λ and μ and self adjoint as linear functions of μ.

Proof. On differentiating

$$[x(\bar{x}; \nu), \lambda] = [\nu, x^*(\bar{x}; \lambda)] \tag{2.6}$$

we obtain

$$[x(\bar{x}; \nu; \mu), \lambda] = [\nu, \underset{\mu}{d_\sigma^{\bar{x}}}\, x^*(\sigma; \lambda)]. \tag{2.7}$$

Clearly

$$[x(\bar{x}; \nu; \mu), \lambda] = [\nu, x^*(\bar{x}; \mu; \lambda)]. \tag{2.8}$$

Hence from (2.7), (2.8) and the positive definiteness of the inner product there results (2.4).

From (2.7) and the definition of $x^*(\bar{x}, \lambda, \nu)$ we have

$$[x(\bar{x}; \nu; \mu), \lambda] = [\mu, x^*(\bar{x}, \lambda, \nu)]. \tag{2.9}$$

But $x(\bar{x}; \nu; \mu)$ is symmetric in ν and μ so that (2.9) makes clear that

[6] Michal (2); Michal and Elconin (1).

$x^*(\bar{x}, \lambda, \nu)$ is self adjoint as a linear function of ν. Hence (2.5) is valid. Finally the bilinearity of $x^*(\bar{x}; \mu; \lambda)$ and $x^*(\bar{x}, \lambda, \mu)$ in λ and μ follows from the continuity of $\underset{\mu}{\overset{\bar{x}}{d_\sigma}} x^*(\sigma; \lambda)$ in $\bar{x}$ and a theorem of Banach on the linearity of the limit of a sequence of linear functions.[7]

THEOREM 2.2. *A necessary and sufficient condition that* $[\xi(x), \eta(x)]$ *be a scalar invariant for an arbitrary contravariant vector*[8] $\xi(x)$ *is that* $\eta(x)$ *be a covariant vector.*[8]

DEFINITION 2.2. *A covariant linear connection is a geometric object with components* $L(x, \eta(x), \xi(x))$ *that are bilinear functions of a covariant vector* $\eta(x)$ *and a contravariant vector* $\xi(x)$ *and such that in the intersection of the geometrical domains of two coördinate systems* $x(P)$ *and* $\bar{x}(P)$, *the components have the law of transformation*

$$(2.10) \quad \bar{L}(\bar{x}, \bar{\eta}(\bar{x}), \bar{\xi}(\bar{x})) = x^*(\bar{x}; L(x, \eta(x), \xi(x)) + x^*(\bar{x}; \bar{\xi}(\bar{x}); \eta(x))$$

under the transformation of coördinates $\bar{x} = \bar{x}(x)$.

Let the covariant vector $\eta(x)$ have a continuous differential $\eta(x; \delta x)$. Then from known theorems on Fréchet differentials and from Theorem 2.1 we obtain

$$(2.11) \qquad \bar{\eta}(\bar{x}; \delta\bar{x}) = x^*(\bar{x}; \eta(x; \delta x)) + x^*(\bar{x}; \delta\bar{x}; \eta(x)).$$

Hence with the aid of (2.10) we obtain

$$(2.12) \quad \bar{\eta}(\bar{x}; \delta\bar{x}) - \bar{L}(\bar{x}, \bar{\eta}(\bar{x}), \delta\bar{x}) = x^*\big(\bar{x}; \eta(x; \delta x) - L(x, \eta(x), \delta x)\big).$$

The steps are reversible so that we have proved the

THEOREM 2.3. *A necessary and sufficient condition that*

$$(2.13) \qquad \eta(x; \delta x) - L(x, \eta(x), \delta x)$$

be a covariant vector whenever $\eta(x)$ *is any continuously differentiable covariant vector is that* $L(x, \eta(x), \delta x)$ *be a covariant linear connection.*

---

[7] Michal (2).

[8] We use contravariant vector and covariant vector as abbreviations for contravariant vector field and covariant vector field respectively. One can, however, define a covariant vector (strict) and contravariant vector (differential) as usual and recast the definitions and theorems (except Theorem 2.3) in the obvious way by substituting contravariant (covariant) vectors for contravariant (covariant) vector fields in the arguments of the linear connections and multilinear forms.

DEFINITION 2.3. *If $L(x, \eta(x), \delta x)$ is a covariant linear connection, then the linear form $\eta(x/\delta x)$ in δx defined by*

$$\eta(x/\delta x) = \eta(x; \delta x) - L(x, \eta(x), \delta x) \tag{2.14}$$

will be called the covariant differential (based on L) of the covariant vector $\eta(x)$.

Let $\Gamma(x, \xi_1(x), \xi_2(x))$ (not necessarily symmetric) be a linear connection,[9] where $\xi_1(x), \xi_2(x)$ are contravariant vectors. This is to be distinguished from the covariant linear connection of Theorem 2.3. The law of transformation[10] for a linear connection is

$$\bar{\Gamma}(\bar{x}, \bar{\xi}_1(\bar{x}), \bar{\xi}_2(\bar{x})) = \bar{x}(x; \Gamma(x, \xi_1(x), \xi_2(x))) + \bar{x}(x; x(\bar{x}; \bar{\xi}_1(\bar{x}); \bar{\xi}_2(\bar{x}))). \tag{2.15}$$

An equivalent law of transformation to (2.15) is

$$\bar{\Gamma}(\bar{x}, \bar{\xi}_1(\bar{x}), \bar{\xi}_2(\bar{x})) = \bar{x}(x; \Gamma(x, \xi_1(x), \xi_2(x))) - \bar{x}(x; \xi_1(x); \xi_2(x)). \tag{2.16}$$

We shall use these laws of transformation in the next section.

3. Covariant differential of multilinear forms in covariant and contravariant vector fields. Since the covariant differential of a covariant vector is a covariant vector depending linearly on an arbitrary contravariant vector, it is clear that the theory of successive covariant differentials can be brought under the theory of covariant differentials of multilinear forms. We shall prove

THEOREM 3.1. *If*

(i) *$F(x, \xi_1(x), \cdots, \xi_r(x), \eta_1(x), \cdots, \eta_s(x))$ is a covariant vector valued multilinear form in the continuously differentiable contravariant vectors $\xi_1(x), \cdots, \xi_r(x)$ and covariant vectors $\eta_1(x), \cdots, \eta_s(x)$,*

(ii) *the partial differential $F(x, \xi_1, \cdots, \xi_r, \eta_1, \cdots, \eta_s; \delta x)$ exists continuous in x,*

then the function $F(x, \xi_1(x), \cdots, \xi_r(x), \eta_1(x), \cdots, \eta_s(x)/\delta x)$ defined by

$$\left\{\begin{aligned} &F(x, \xi_1(x), \cdots, \xi_r(x), \eta_1(x), \cdots, \eta_s(x)/\delta x) \\ &\quad = F(x, \xi_1(x), \cdots, \xi_r(x), \eta_1(x), \cdots, \eta_s(x); \delta x) \\ &\qquad - \sum_{i=1}^{r} F(x, \xi_1(x), \cdots, \xi_{i-1}(x), \Gamma(x, \xi_i(x), \delta x), \\ &\qquad\qquad\qquad \xi_{i+1}(x), \cdots, \xi_r(x), \eta_1(x), \cdots, \eta_s(x)) \\ &\qquad + \sum_{i=1}^{s} F(x, \xi_1(x), \cdots, \xi_r(x), \eta_1(x), \cdots, \eta_{i-1}(x), \\ &\qquad\qquad\qquad L(x, \eta_i(x), \delta x), \eta_{i+1}(x), \cdots, \eta_s(x)) \\ &\qquad - L(x, F(x, \xi_1(x), \cdots, \xi_r(x), \eta_1(x), \cdots, \eta_s(x)), \delta x) \end{aligned}\right. \tag{3.1}$$

---

[9] Michal (1), (2).

[10] Michal (1).

is a covariant vector valued multilinear form in $\xi_1, \cdots, \xi_r, \eta_1, \cdots, \eta_s, \delta x$. *We shall call* $F(x, \xi_1, \cdots, \xi_r, \eta_1, \cdots, \eta_s/\delta x)$ *the covariant differential of* $F(x, \xi_1, \cdots, \xi_r, \eta_1, \cdots, \eta_s)$.

Proof. We shall give the details of proof for $r = 1$, $s = 1$ as the proof for the general case, although lengthy, differs in no essential manner from that of this special case.

By hypothesis

$$\bar{F}\big(\bar{x}, \bar{\xi}(\bar{x}), \bar{\eta}(\bar{x})\big) = x^*\big(\bar{x}; F(x, \xi(x), \eta(x))\big) \tag{3.2}$$

from which we obtain with the aid of Theorem 2.1

$$\left\{\begin{aligned} \bar{F}\big(\bar{x}, \bar{\xi}(\bar{x}), \bar{\eta}(\bar{x}); \delta\bar{x}\big) = {} & x^*\big(\bar{x}; F(x, \xi(x), \eta(x); \delta x)\big) \\ & + x^*\big(\bar{x}; \delta\bar{x}; F(x, \xi(x), \eta(x))\big) \\ & + x^*\big(\bar{x}; d^{x}_{\delta x}{}^{\sigma} F(x, \xi(\sigma), \eta(\sigma))\big) - d^{\bar{x}}_{\delta\bar{x}}{}^{\sigma} \bar{F}(\bar{x}, \bar{\xi}(\sigma), \bar{\eta}(\sigma)). \end{aligned}\right. \tag{3.3}$$

On using (2.16) and (3.2) we find that

$$\left\{\begin{aligned} \bar{F}\big(\bar{x}, \bar{\Gamma}(\bar{x}, \bar{\xi}(\bar{x}), \delta\bar{x}), \bar{\eta}(\bar{x})\big) = {} & x^*\big(\bar{x}; F(x, \Gamma(x, \xi(x), \delta x), \eta(x))\big) \\ & - \bar{F}\big(\bar{x}, \bar{x}(x; \xi(x); \delta x), \bar{\eta}(\bar{x})\big). \end{aligned}\right. \tag{3.4}$$

Similarly from (2.10) and (3.2)

$$\left\{\begin{aligned} \bar{F}\big(\bar{x}, \bar{\xi}(\bar{x}), \bar{L}(\bar{x}, \bar{\eta}(\bar{x}), \delta\bar{x})\big) = {} & x^*\big(\bar{x}; F(x, \xi(x), L(x, \eta(x), \delta x))\big) \\ & + \bar{F}\big(\bar{x}, \bar{\xi}(\bar{x}), x^*(\bar{x}; \delta\bar{x}; \eta(x))\big). \end{aligned}\right. \tag{3.5}$$

Evidently

$$\left\{\begin{aligned} \bar{L}\big(\bar{x}, \bar{F}(\bar{x}, \bar{\xi}(\bar{x}), \bar{\eta}(\bar{x})), \delta\bar{x}\big) = {} & x^*\big(\bar{x}; L(x, F(x, \xi(x), \eta(x)), \delta x)\big) \\ & + x^*\big(\bar{x}; \delta\bar{x}; F(x, \xi(x), \eta(x))\big). \end{aligned}\right. \tag{3.6}$$

Taking the differential of (2.1) and using (3.2) we obtain

$$\left\{\begin{aligned} \bar{F}\big(\bar{x}, \delta\bar{\xi}(\bar{x}), \bar{\eta}(\bar{x})\big) = {} & \bar{F}\big(\bar{x}, \bar{x}(x; \xi(x); \delta x), \bar{\eta}(\bar{x})\big) \\ & + x^*\big(\bar{x}; F(x, \delta\xi(x), \eta(x))\big). \end{aligned}\right. \tag{3.7}$$

Similarly from (2.11) and (3.2)

$$\left\{\begin{aligned} \bar{F}\big(\bar{x}, \bar{\xi}(\bar{x}), \delta\bar{\eta}(\bar{x})\big) = {} & \bar{F}\big(\bar{x}, \bar{\xi}(\bar{x}), x^*(\bar{x}; \delta\bar{x}; \eta(x))\big) \\ & + x^*\big(\bar{x}; F(x, \xi(x), \delta\eta(x))\big). \end{aligned}\right. \tag{3.8}$$

Reducing (3.3) by means of (3.7) and (3.8)

$$
(3.9) \begin{cases} \bar{F}\left(\bar{x}, \bar{\xi}(\bar{x}), \bar{\eta}(\bar{x}) ; \delta \bar{x}\right) \\ \quad = x^*\left(\bar{x} ; F(x, \xi(x), \eta(x) ; \delta x)\right) + x^*\left(\bar{x} ; \delta \bar{x} ; F(x, \xi(x), \eta(x))\right) \\ \quad\quad - \bar{F}\left(\bar{x}, \bar{x}(x ; \xi(x) ; \delta x), \bar{\eta}(\bar{x})\right) - \bar{F}\left(\bar{x}, \bar{\xi}(\bar{x}), x^*(\bar{x} ; \delta \bar{x} ; \eta(x))\right). \end{cases}
$$

Finally with the aid of (3.4), (3.5), (3.6), and (3.9) we obtain

$$
\bar{F}\left(\bar{x}, \bar{\xi}(\bar{x}), \bar{\eta}(\bar{x}) / \delta \bar{x}\right) = x^*\left(\bar{x} ; F(x, \xi(x), \eta(x) / \delta x)\right),
$$

which completes the proof of the special case $r = 1$, $s = 1$.

The following two theorems can now be proved without much difficulty.

THEOREM 3.2. *If in the hypotheses of Theorem 3.1, the function* $F(x, \xi_1, \cdots, \xi_r, \eta_1, \cdots, \eta_s)$ *is taken to be a contravariant vector valued multilinear form, then the function* $F(x, \xi_1, \cdots, \xi_r, \eta_1, \cdots, \eta_s / \delta x)$ *defined by*

$$
(3.10) \begin{cases} F\left(x, \xi_1(x), \cdots, \xi_r(x), \eta_1(x), \cdots, \eta_s(x) / \delta x\right) \\ \quad = F\left(x, \xi_1(x), \cdots, \xi_r(x), \eta_1(x), \cdots, \eta_s(x) ; \delta x\right) \\ \quad\quad - \sum_{i=1}^{r} F\left(x, \xi_1, \cdots, \xi_{i-1}, \Gamma(x, \xi_i(x), \delta x), \xi_{i+1}, \cdots, \xi_r, \eta_1, \cdots, \eta_s\right) \\ \quad\quad + \sum_{i=1}^{s} F\left(x, \xi_1, \cdots, \xi_r, \eta_1, \cdots, \eta_{i-1}, L(x, \eta_i, \delta x), \eta_{i+1}, \cdots, \eta_s\right) \\ \quad\quad + \Gamma\left(x, F(x, \xi_1, \cdots, \xi_r, \eta_1, \cdots, \eta_s), \delta x\right) \end{cases}
$$

is a contravariant valued multilinear form in $\xi_1, \cdots, \xi_r, \eta_1, \cdots, \eta_s, \delta x$. *We shall call* $F(x, \xi_1, \cdots, \xi_r, \eta_1, \cdots, \eta_s / \delta x)$ *the covariant differential of* $F(x, \xi_1, \cdots, \xi_r, \eta_1, \cdots, \eta_s)$.

THEOREM 3.3. *If in the hypotheses of Theorem 3.1, the function* $F(x, \xi_1, \cdots, \xi_r, \eta_1, \cdots, \eta_s)$ *is taken to be an absolute scalar multilinear form (with numerical values or with values in a Banach space), then the function* $F(x, \xi_1, \cdots, \xi_r, \eta_1, \cdots, \eta_s / \delta x)$ *defined by*

$$
(3.11) \begin{cases} F\left(x, \xi_1(x), \cdots, \xi_r(x), \eta_1(x), \cdots, \eta_s(x) / \delta x\right) \\ \quad = F(x, \xi_1, \cdots, \xi_r, \eta_1, \cdots, \eta_s ; \delta x) \\ \quad\quad - \sum_{i=1}^{r} F\left(x, \xi_1, \cdots, \xi_{i-1}, \Gamma(x, \xi_i, \delta x), \xi_{i+1}, \cdots, \xi_r, \eta_1, \cdots, \eta_s\right) \\ \quad\quad + \sum_{i=1}^{s} F\left(x, \xi_1, \cdots, \xi_r, \eta_1, \cdots, \eta_{i-1}, L(x, \eta_i, \delta x), \eta_{i+1}, \cdots, \eta_s\right) \end{cases}
$$

is an absolute scalar multilinear form (with numerical or Banach values) in $\xi_1, \cdots, \xi_r, \eta_1, \cdots, \eta_s, \delta x$. *We shall call* $F(x, \xi_1, \cdots, \xi_r, \eta_1, \cdots, \eta_s / \delta x)$ *the covariant differential of* $F(x, \xi_1, \cdots, \xi_r, \eta_1, \cdots, \eta_s)$.

To have the successive covariant differential $\eta(x/\delta_1 x/\cdots/\delta_p x)$ well defined in all coördinate systems, it is sufficient to assume that (1) the covariant vector $\eta(x)$ possesses a continuous p-th differential; (2) $L(x,\eta,\delta x)$ has a continuous p-th partial differential in the first place; (3) $\Gamma(x,\xi_1,\xi_2)$ has a continuous $(p-1)$-st partial differential in the first place; (4) $\bar{x}(x)$ has a continuous $(p+2)$-nd differential; and (5) $x^*(\bar{x};\lambda)$ has a continuous $(p+1)$-st differential in $\bar{x}$.

By calculation we obtain the commutation rule

$$
\begin{aligned}
(3.12)\quad & \eta(x/\delta_1 x/\delta_2 x) - \eta(x/\delta_2 x/\delta_1 x) \\
& = - L(x,\eta(x),\delta_1 x,\delta_2 x) - 2\eta(x/\Omega(x,\delta_1 x,\delta_2 x)),
\end{aligned}
$$

where

$$
(3.13)\quad \left\{ \begin{aligned} L(x,\eta(x),\delta_1 x,\delta_2 x) &= L(x,\eta(x),\delta_1 x;\delta_2 x) - L(x,\eta(x),\delta_2 x;\delta_1 x) \\ &\quad + L\big(x,L(x,\eta(x),\delta_2 x),\delta_1 x\big) - L\big(x,L(x,\eta(x),\delta_1 x),\delta_2 x\big) \end{aligned} \right.
$$

and

$$
(3.14)\qquad \Omega(x,\delta_1 x,\delta_2 x) = \tfrac{1}{2}\{\Gamma(x,\delta_1 x,\delta_2 x) - \Gamma(x,\delta_2 x,\delta_1 x)\}.
$$

Since $\Omega(x,\delta_1 x,\delta_2 x)$ is the contravariant vector valued *torsion* form, it follows from (3.12) and Theorem 3.1 that the trilinear form $L(x,\eta(x),\delta_1 x,\delta_2 x)$ is a covariant vector valued trilinear form, called the (covariant vector valued) *curvature form.*

Suppose now that $F(x,\eta(x),\delta x)$ is bilinear in covariant vectors $\eta(x)$ and in δx, and suppose further that $F(x,\eta(x),\delta x)$ is the component of a geometric object. On making special use of the properties of the adjoints and the law of transformation of the linear connection one can demonstrate without much difficulty the following theorem.

THEOREM 3.4. *A necessary and sufficient condition that the adjointness relation*

$$
[\Gamma(x,\xi(x),\delta x),\eta(x)] = [\xi(x),F(x,\eta(x),\delta x)]
$$

be a geometric condition (i. e., continues to hold under a transformation of coördinates) is that $F(x,\eta(x),\delta x)$ be a covariant linear connection.

The importance of a relation of the above type between the linear connection Γ and the covariant linear connection L is made clear from the following result.

THEOREM 3.5. *A necessary and sufficient condition that*

$$
\delta[\xi(x),\eta(x)] = [\xi(x/\delta x),\eta(x)] + [\xi(x),\eta(x/\delta x)]
$$

for all continuously differentiable contravariant vectors $\xi(x)$ *and covariant vectors* $\eta(x)$ *is that the covariant linear connection* $L(x, \eta(x), \delta x)$ *be the adjoint of the linear connection* $\Gamma(x, \xi(x), \delta x)$ *considered as a linear function of* $\xi(x)$.

CALIFORNIA INSTITUTE OF TECHNOLOGY,
PASADENA, CALIFORNIA.

REFERENCES.

A. D. Michal, I. E. Highberg and A. E. Taylor:

(1) "Abstract Euclidean spaces with independently postulated analytical and geometrical metrics," *Annali di Pisa* (in press).

A. D. Michal:

(1) "General tensor analysis," *Bulletin of the American Mathematical Society* (in press).

(2) "Postulates for a linear connection," *Annali di Matematica* (in press).

A. D. Michal and V. Elconin:

(1) "Completely integrable differential equations in abstract spaces," *Acta Mathematica* (in press).

A TYPE OF HOMOGENEITY FOR CONTINUOUS CURVES.[1]

By Charles H. Wheeler, III.

Introduction. A set of points M is said to be *homogeneous*[2] if, given any two points x and y of M, it is possible to find a homeomorphism which will send M into itself in such a way that x is sent into y.

A set of points M is said to be *bi-homogeneous*[3] if given any two points x and y of M there exists a homeomorphism which sends M into itself in such a way that x is sent into y and y is sent into x.

We will investigate the conditions under which a compact, locally connected continuum may be *cyclic element homogeneous*, i. e., given any two true cyclic elements[4] of a set M, there exists a homeomorphism which sends M into itself in such a way that one of the given true cyclic elements is sent into the other. A compact locally connected continuum is a continuous curve in the sense that it is a set of points which is the image of the unit interval under a continuous transformation.

The case where M contains only a finite number of true cyclic elements will be completely treated, while some results will be stated for the case where M contains infinitely many true cyclic elements.

Two simple closed curves joined by a simple arc provide an example of a cyclic element homogeneous set. The curve illustrated in Fig. 1 is not cyclic element homogeneous because the true cyclic element marked C_1 cannot be sent into any of the other true cyclic elements by a homeomorphism which sends the set into itself. In Fig. 2, also in Fig. 1, each true cyclic element is homeomorphic with each of the remaining ones, but in Fig. 2 the cyclic element marked C_2 cannot be sent into any of the others by a homeomorphism which sends the set into itself. Fig. 3 is an example of a set of points which contains an infinite number of true cyclic elements and is cyclic element homogeneous.

---

[1] Received April 18, 1935; revised October 15, 1936.

[2] See Kuratowski, *Fundamenta Mathematicae*, T. 3 (1922), pp. 14-19, also Mazurkiewicz, *Fundamenta Mathematicae*, T. 5 (1924), pp. 137-146.

[3] Kuratowski, *loc. cit.*

[4] The cyclic elements of a locally connected continuum are (1) all cut points of the continuum and (2) the set of all points conjugate to a point p, where p is any non-cut point of the continuum. A true cyclic element is one which does not reduce to a single point. See G. T. Whyburn, "Concerning the structure of a continuous curve," *American Journal of Mathematics*, vol. 50 (1928), pp. 167-194, and Kuratowski and Whyburn, "Sur les eléménts cycliques et leurs applications," *Fundamenta Mathematicae*, T. 16 (1930), pp. 305-331.

1. *Preliminary theorems.* Let M be any compact and locally connected continuum, which we shall consider as a space. Designate by H the smallest A-set [5] which contains all the true cyclic elements C_i in M, i. e., an A-set containing all the true cyclic elements C_i and not a proper subset of any A-set

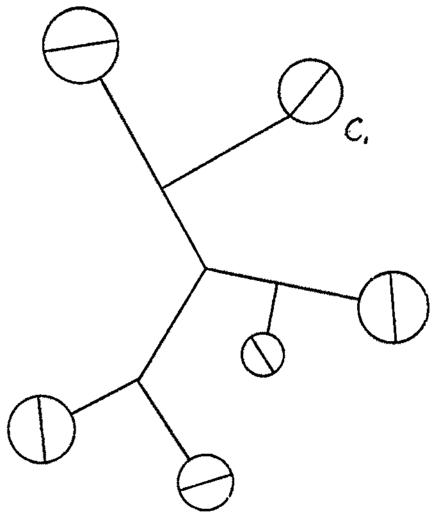


Fig. 1

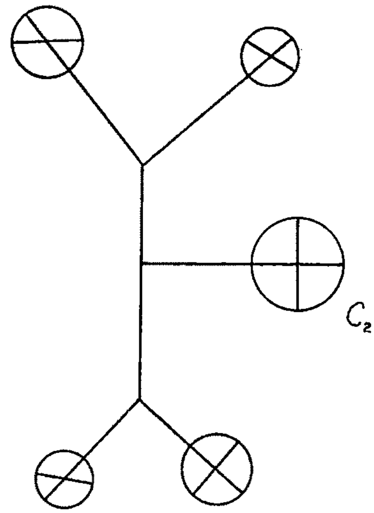


Fig. 2

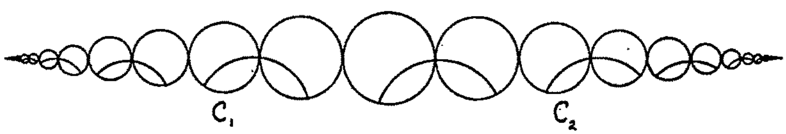


Fig. 3

containing all the true cyclic elements. The set H may be obtained by taking the product of all A-sets in M which contain all the true cyclic elements. Then since [6] the product of any number of A-sets is an A-set, it follows H is an

[5] A closed set which has the property that if $x, y \in A$ then every arc xy in the space is contained in A.

[6] Cf. Kuratowski and Whyburn, *loc. cit.*, Theorem 4: 1.

A-set and clearly it is the smallest such set containing all the true cyclic elements. If there are only a finite number of true cyclic elements, the set H may be obtained by choosing a non-cut point p_i from each true cyclic element C_i, and forming all the possible pairs of these points. Then for each pair p_i, p_j form the cyclic chain $C_k(p_i, p_j)$. We then have $H = \sum_1^n C_k(p_i, p_j)$.

1.1. THEOREM. *If M contains only a finite number (> 1) of true cyclic elements, then there exists at least two true cyclic elements each of which has only one point in common with the closure of the remainder of H.*

Proof. This follows immediately from a theorem of G. T. Whyburn.[7] He proves that if a continuum has more than one cyclic element then it contains at least two nodes, where a node is defined as an end point or a true cyclic element which contains only one cut point. Hence H must contain at least two true cyclic elements which contain only one cut point since it contains no end points.

1.2. THEOREM. *If T is any homeomorphism which sends M into itself, then $T(H) = H$.*

Proof. The set H is uniquely defined as the smallest A-set containing all the true cyclic elements of M. Any homeomorphism T will send H into an A-set which contains all the true cyclic elements of M, thus $H \subset T(H)$.

Since T^{-1} is a homeomorphism, $H \subset T^{-1}(H)$. Operating upon this with T we have $T(H) \subset TT^{-1}(H) = H$. Thus $T(H) \subset H$. It follows from the above two inclusions that $T(H) = H$, and the theorem is proved.

We will consider H to be our space for the remainder of this paper. H is a compact locally connected continuum.

1.3. *Definition.* A set H is said to be *cyclic element homogeneous* if, given any two true cyclic elements of H, then there exists a homeomorphism T such that $T(H) = H$ and one of the given true cyclic elements is sent into the other.

From this definition we have immediately the following:

1.4. THEOREM. *If H is cyclic element homogeneous, then each true cyclic element contains the same number (finite or infinite) of cut points of H.*

2. *The case where H contains only a finite number of true cyclic elements.* If the set H contains only a finite number (> 1) of true cyclic elements,

[7] "Concerning the structure of a continuous curve," *loc. cit.*, p. 180.

there exists a true cyclic element C_s which contains only one point which cuts H by 1. 1. Then every true cyclic element of H has only one point in common with the closure of the remainder of H, by 1. 3, i. e., $C_i \cdot \overline{H - C_i} =$ a single point.

Definition. Let $p_1{}^i$ be the point of C_i such that $p_1{}^i \in \overline{H - C_i}$. The sum of all true cyclic elements which contain $p_1{}^i$ is called a *cluster* of true cyclic elements. The point $p_1{}^i = p_j$ is called the center of the cluster K_j.

Then $K_j = \sum_{i=1}^{n} C_i{}^j$ and $\prod_{i=1}^{n} C_i{}^j = p_j$. The centers of every two clusters are joined by a simple arc in H, since H is an A-set. The arc $p_j p_k$ is such that

$$p_j p_k \cdot K_j = p_j \quad \text{and} \quad p_j p_k \cdot K_k = p_k,$$

for we have seen above that no $C_i{}^j$ of K_j can have more than one point in common with $\overline{H - C_i{}^j}$.

2. 1. Theorem. *If H is cyclic element homogeneous and contains only a finite number of true cyclic elements, then there is the same number of true cyclic elements in each cluster.*

Proof. Suppose the contrary, then there exists some two clusters such that

$$K_1 = \sum_1^n C_i{}^1, \qquad K_2 = \sum_1^m C_i{}^2, \qquad n > m.$$

The points p_1 and p_2 are the centers of the clusters K_1 and K_2 respectively. Since H is cyclic element homogeneous, let $T(H) = H$ in such a way that $T(C_1{}^1) = C_1{}^2$. Now $C_1{}^1 \cdot \overline{K_1 - C_1{}^1} = p_1$ and $C_1{}^2 \cdot \overline{K_2 - C_1{}^2} = p_2$, and hence $T(p_1) = p_2$.

Since $C_2{}^1 \supset p_1$, then $T(C_2{}^1) =$ some $C_i{}^2$ for some i, $i = 2, 3, \cdots, m$. This is also true for the remaining $n - 2$ true cyclic elements in K_1, but there are $m - n$ less true cyclic elements in K_2 than in K_1, so this is impossible. Thus the supposition that the clusters do not contain the same number of true cyclic elements leads to a contradiction, and the theorem is proved.

2. 2. Corollary. *If a true cyclic element of one cluster is transformed by a homeomorphism T into a true cyclic element of another cluster, then the first cluster is transformed into the second cluster by T.*

2. 3. Theorem. *If H is cyclic element homogeneous and contains only a finite number of true cyclic elements, then no cluster cuts H.*

Proof. Suppose there exists no cluster which does not cut H. There are only a finite number of clusters since there are only a finite number of true cyclic elements in H. Then $H - K_1 = H_1^1 + H_2^1$, where

$$H_1^1 \cdot \overline{H_2^1} = \overline{H_1^1} \cdot H_2^1 = 0 \quad \text{and} \quad H_1^1 \neq 0 \neq H_2^1.$$

H_1^1 contains at least one cluster K_{i_1}. Let K_{n_2} be the cluster in H_1^1 with the least subscript. Then $H - K_{n_2} = H_1^2 + H_2^2$, where

$$H_1^2 \cdot \overline{H_2^2} = \overline{H_1^2} \cdot H_2^2 = 0, \quad H_1^2 \neq 0 \neq H_2^2 \quad \text{and} \quad H_2^2 \supset K_1.$$

H_1^2 contains at least one cluster K_{i_2}, $K_{i_2} \neq K_{n_2}$. Let K_{n_3} be the cluster in H_1^2 with the least subscript. Then $H - K_{n_3} = H_1^3 + H_2^3$, where

$$H_1^3 \cdot \overline{H_2^3} = \overline{H_1^3} \cdot H_2^3 = 0, \quad H_1^3 \neq 0 \neq H_2^3 \quad \text{and} \quad H_2^3 \supset K_{n_2}.$$

H_1^3 contains at least one cluster K_{i_3}, $K_{i_3} \neq K_{n_3}$.

This can be carried on indefinitely, but this is impossible for there are only a finite number of clusters. Thus there exists at least one cluster K_t which does not cut H.

Let K_p be any cluster of H different from K_t. By 2.2 $H - K_t$ is homeomorphic with $H - K_p$ and hence K_p can not cut H. Thus no cluster cuts H.

2.4. *Definitions.*

2.41. A compact locally connected continuum H is said to be *symmetrical* with respect to a cyclic element C (whether C be a true cyclic element or not) if every component of $H - C$ is homeomorphic with every other component of $H - C$. We will call C the center of symmetry.

2.42. A set of points H is said to be *cyclic symmetrical* with respect to a cyclic element C (whether C be a true cyclic element or not) if it is symmetrical and any true cyclic element of one component of $H - C$ may be sent into any true cyclic element of another component of $H - C$ by a homeomorphism which sends the first component into the second.

2.43. A *major branch* at a branch point x is the component of $H - x$ which contains the center of symmetry.

2.44. A *minor branch* at a branch point x is a component of $H - x$ which does not contain the center of symmetry.

2.5. THEOREM. *If H is cyclic element homogeneous and contains only a finite number of true cyclic elements, then there exists a point c such that H is cyclic symmetrical with respect to c.*

Proof. We saw that the centers of every pair of clusters are joined by an arc in H, and that this arc has only the centers of the clusters in common with the clusters. Also by 2.3 no cluster cuts H. Thus $\overline{H - \sum_{1}^{m} K_i}$ is an acyclic curve with a finite number of branches.

There is a center p_i of a cluster K_i at the end of each branch. Form all the possible pairs of the points p_i. Take a pair p_1, p_2 which has a maximum number of branch points on the arc joining them. If this number is odd let the middle branch point be c, if it is even let c be any point on the open arc joining the two middle branch points. The point c is uniquely determined in the case where the maximum number is odd, and uniquely determined to the extent of being any one of the points on an open arc in the case where the maximum number is even. Now

$$H - c = S_1 + S_2 + S_3 + \cdots + S_n,$$

where $S_1, S_2, \cdots, S_n$ are components. If the number of branch points on the arc joining p_1 and p_2 is even there are only two components, while if this number is odd there may be any finite number of components. Let $S_1 \supset p_1$ and $S_2 \supset p_2$.

Number the branch points on S_1 in the following way: Let the first branch point out from c be x^1, then on each minor branch from x^1 let the first branch points be $x_1{}^2, x_2{}^2, \cdots, x_m{}^2$. The branch point on the branch containing p_1 we will call x^2. Let the first branch points on the minor branches from $x_1{}^2$ be $x^3{}_{1,1}, x^3{}_{1,2}, \cdots, x^3{}_{1,k}$. The branch point on the branch containing p_1 we will call x^3. Continue in this manner until all the branch points have been numbered. Number the branch points on S_2 in the same manner except with the use of y instead of x.

The number of branch points from c to p_1 is the same as the number from c to p_2; let us assume this number is k.

We must now show that there exists a homeomorphism T such that $T(S_1) = S_2$. Since H is cyclic element homogeneous, let T be a homeomorphism of H into itself such that $T(C_1{}^1) = C_1{}^2$ where $C_1{}^1 \supset p_1$ and $C_1{}^2 \supset p_2$. Then $T(p_1) = p_2$ and $T(K_1) = K_2$. The arc p_1x^k is sent into the arc p_2y^k and x^k into y^k by T. At the branch points x^k and y^k there are the same number of branches because x^k and y^k correspond to each other under a homeomorphism. There is no branch point on any of the minor branches at x^k or y^k for if there were one on some minor branch, say at y^k, there would be more branch points on the arc joining one of the centers p_i (on this branch) and p_1 than on the arc p_1p_2. Then p_1, p_2 would not be a maximal pair.

The arc x^kx^{k-1} is sent into the arc y^ky^{k-1} by T so that x^{k-1} is sent into y^{k-1}. There are the same number of branches at x^{k-1} as at y^{k-1} because x^{k-1} and y^{k-1} corrrespond to each other under a homeomorphism. There can not be more than one branch point on any of the minor branches at x^{k-1} or y^{k-1}; for if there were, p_1, p_2 would not have been a maximal pair. There must be one branch point on each of the minor branches at x^{k-1} and y^{k-1} and the number of minor branches at these branch points is the same as the number of minor branches at x^k or y^k.

Now continue down the major branch at x^{k-1} and y^{k-1}. The arc $x^{k-1}x^{k-2}$ is sent into the arc $y^{k-1}y^{k-2}$ by T, and then x^{k-2} is sent into y^{k-2}. The number of minor branches at x^{k-2} and y^{k-2} is the same, because x^{k-2} and y^{k-2} correspond to each other under a homeomorphism. The minor branches at x^{k-2} are homeomorphic with the minor branches at y^{k-2}. On continuing in this manner until we reach $T(x^1) = y^1$, it can be shown that the minor branches at x^1 are homeomorphic with the minor branches at y^1. Then the arc $x^1c - c$ is sent into the arc $y^1c - c$ by T. Therefore $T(S_1) = S_2$.

If there are only two components of $H - c$, the theorem is proved. In the case where there are more than two components of $H - c$, take any component S_3 different from S_1 and S_2. Let p_3 be the center of the cluster K_3 such that the number of branch points on the arc joining it to c is a maximum for all centers p_i in S_3. Let the number of branch points on this arc be m. Then $m \leq k$, for if it were greater p_1, p_2 would not have been a maximal pair. Number the branch points on S_3 the same as on S_1, denoting them by z instead of x.

Since H is cyclic element homogeneous, let $T(H) = H$ in such a way that $T(C_1^1) = C_1^3$, where $C_1^1 \subset K_1 \subset S_1$ and $C_1^3 \subset K_3 \subset S_3$. Then $T(K_1) = K_3$, $T(p_1) = p_3$, $T(p_1x^k) = p_3z^m$ and $T(x^k) = z^m$. The number of branches at x^k is the same as the number at z^m, for x^k and z^m correspond to each other under a homeomorphism. There is no branch point on any of the minor branches at x^k or z^m, for if there were, p_3 would not have been a maximal number of branch points from c. Then $T(x^kx^{k-1}) = z^mz^{m-1}$ and $T(x^{k-1}) = z^{m-1}$. The number of branch points at x^{k-1} and z^{m-1} are the same since x^{k-1} and z^{m-1} correspond under a homeomorphism. The minor branches at z^{m-1} are homeomorphic and homeomorphic to the minor branches at x^{k-1}.

It is thus seen that if $m = k$, S_1 is homeomorphic with S_3. If $m < k$, say $m + 1 = k$, then one minor branch at x^1 is sent into S_3 and x^1 is sent into c. Since x^1 and c correspond to each other under a homeomorphism, there are the same number of branches at x^1 as there are at c. This may or may not be true; if not, it is seen immediately that $T(H) \neq H$, and therefore

$m = k$. If the number of branches at x^1 and c are the same, consider the number of branch points on the arc from x^1 to $p_2 \in S_2$. There are $k+1$. The branch from x^1 containing these must be sent into a component of $H - c$ different from S_3; but we have seen that the maximum number of branch points from c to the center of any cluster in $H - c$ was k. Therefore this is impossible and $T(H) \neq H$. Thus $m = k$. Therefore all of the components of $H - c$ are homeomorphic.

It must now be shown that any true cyclic element in one component may be sent into any true cyclic element in another component by a homeomorphism which sends the first component into the second. Let $C_1{}^i$ be any true cyclic element in S_r and $C_1{}^j$ any in S_t. Now $C_1{}^i \subset K_i$ and $C_1{}^j \subset K_j$. Since H is cyclic element homogeneous, let $T(H) = H$ in such a manner that $T(C_1{}^i) = C_1{}^j$. It has already been shown that S_r and S_t are homeomorphic, that every cluster on S_r and S_t are the same number (k) of branch points away from c, that at each i-th branch point out from c $(i = 1, 2, 3, \cdots, k)$ there are the same number of minor branches which are homeomorphic with one another. Let the branch points on S_r and S_t be numbered in the same manner as were the branch points on S_1 and denoted by u and v, respectively. We have then $T(K_i) = K_j$, $T(p_i) = p_j$, $T(p_i u^k) = p_j v^k$ and $T(u^k) = v^k$. The remaining minor branches at u^k are sent into the remaining minor branches at v^k by T. Then $T(u^k u^{k-1}) = v^k v^{k-1}$ and $T(u^{k-1}) = v^{k-1}$. The remaining minor branches at u^{k-1} are sent into the remaining minor branches at v^{k-1} by T. Continuing this finally yields $T(u^1) = v^1$. The remaining minor branches at u^1 are sent into the remaining minor branches at v^1. Then $T(u^1 c - c) = v^1 c - c$, thus $T(S_r) = S_t$. Q. E. D.

2. 6. *Summary.* We have proved that if H is cyclic element homogeneous and contains only a finite number of true cyclic elements C_i then

1) $C_i \cdot \overline{H - C_i}$ = a single point, for each i.
2) Each cluster K_i has the same number of true cyclic elements.
3) $H - K_i$, for every i, is connected.
4) $\overline{H - \sum_1^n K_i}$ is an acyclic curve with a finite number of branches.
5) There exists a point c such that
   a. Every cluster is the same number of branch points distant from c.
   b. At each i-th branch point from c, $(i = 1, 2, \cdots, k)$ there are the same number of branches, and all the minor branches are homeomorphic.
   c. On every component of $H - c$ there are the same number of branch points and the same number of clusters.

d. Every component of $H - c$ is homeomorphic with every other and any true cyclic element of one component may be sent into any true cyclic element of any other component by a homeomorphism T which sends the first component into the second and which sends H into itself.

6) All the true cyclic elements are homeomorphic, and if $p_i^r \epsilon C_r$ and $p_i^s \epsilon C_s$ such that $C_r \cdot \overline{H - C_r} \supset p_i^r$ and $C_s \cdot \overline{H - C_s} \supset p_i^s$, then the homeomorphism $W(C_r) = C_s$ is such that $W(p_i^r) = p_i^s$ for some i.

Fig. 4 is an example of a set L which is cyclic element homogeneous. The point c is the center of symmetry, and there are three components of $L - c$.

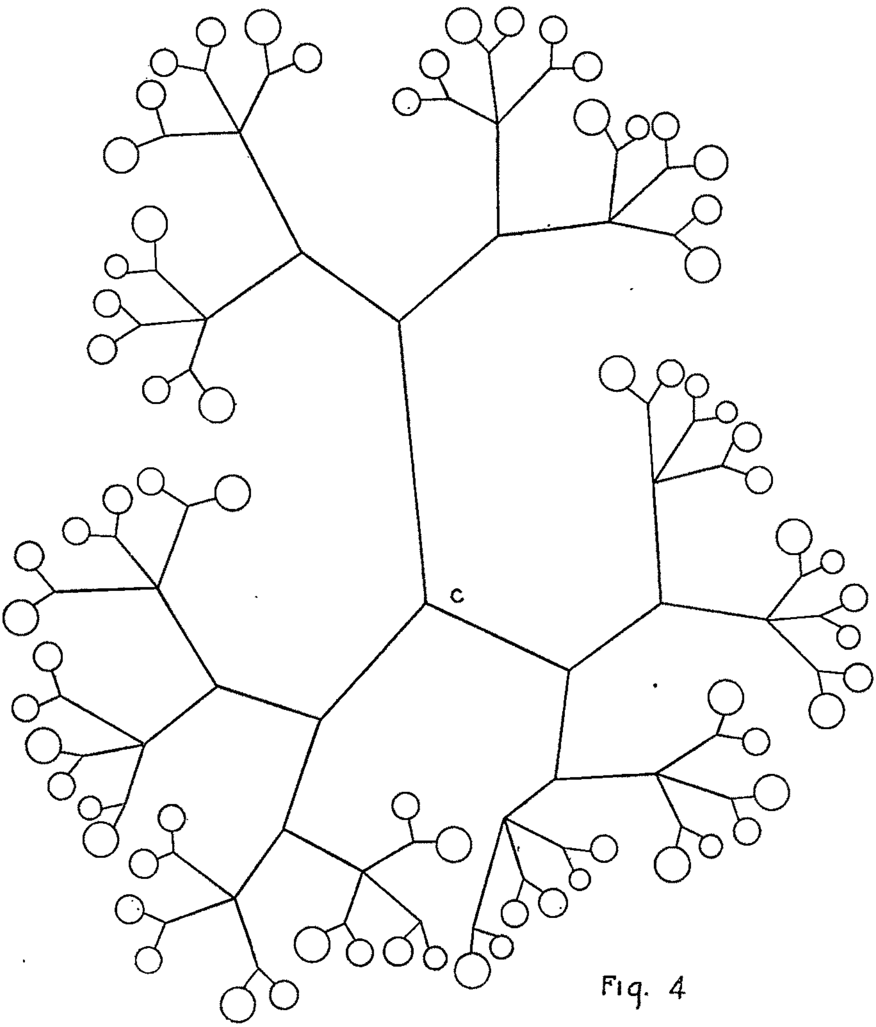


Fig. 4

2.7. THEOREM. *If H contains only a finite number of true cyclic elements, in order that H be cyclic element homogeneous it is necessary and sufficient that* (1) *the true cyclic elements be grouped in clusters with the same number in each cluster,* (2) *no cluster cuts H,* (3) *there exists a point c of H such that each cluster is the same number, k, of branch points away from c,* (4) *at each i-th branch point $(i = 1, 2, \cdots, k)$ from c there are the same number of branches, and* (5) *if C_r and C_s are any two true cyclic elements of H, there exists a homeomorphism of C_r into C_s which sends the cut points of H on C_r into the cut points of H on C_s.*

Proof. The necessity follows from 2.5. To show the sufficiency take any two components S_1, S_2 of $H - c$. Denote the branch points on S_1 by x and those on S_2 by y, as was done in 2.5. Go out along the arcs from c to x^1 and to y^1 lying in S_1 and S_2 respectively. There are the same number of branches at x^1 as at y^1, by hypothesis. Out from x^1 and y^1 on each minor branch there is a second branch point from c. There are the same number of branches at each of these by hypothesis. Out from the second branch points from c on each of the minor branches is the third branch point from c, and there are the same number of branches at each of these.

Continue this until we get to the k-th branch points from c. By hypothesis there are the same number of branches at each of these points. Out from the k-th branch points on each of the minor branches is a cluster of true cyclic elements, for otherwise this branch would not have been in H. There can not be any branch point on any of these branches, for if there were there would be at least two clusters which had more than k branch points on the arc from the center of the cluster to c. There is no cluster on any minor branch from x^1 and y^1 before the k-th branch point, for if there were there would be less than k branch points on the arc from its center to c. Thus there is a cluster of true cyclic elements at the end of each minor branch from the k-th branch points. By hypothesis there is the same number of true cyclic elements in each cluster. Since no cluster cuts H, every two of the clusters are joined by an arc in H which passes through at least one branch point. Each true cyclic element contains only one point which cuts H, namely the center of the cluster in which the true cyclic element is contained. This follows from the fact that no cluster cuts H.

It must now be shown that there exists a homeomorphism T such that if there be given any two true cyclic elements C_i, C_j of H, then $T(H) = H$ in such a way that $T(C_i) = C_j$. Take any two true cyclic elements C_1 and C_2. By hypothesis there exists a homeomorphism W such that $W(C_1) = C_2$ and $W(p_1) = p_2$ where $p_1 \epsilon C_1 \cdot \overline{H - C_1}$ and $p_2 \epsilon C_2 \cdot \overline{H - C_2}$. Define the homeo-

morphism $T = W$ over C_1. There are the same number of true cyclic elements in each cluster, hence the definition of T can be extended so that $T(K_1) = K_2$ where $K_1 \supset C_1$ and $K_2 \supset C_2$. If p_1x^k and p_2y^k are the arcs from K_1 and K_2 to the branch points x^k and y^k respectively, we so extend the definition of T that $T(p_1x^k) = p_2y^k$, and $T(x^k) = y^k$; also $T(b^i{}_{x,k}) = b^i{}_{y,k}$ where $b^i{}_{x,k}$ are the minor branches at x^k which do not contain p_1 and $b^i{}_{y,k}$ the minor branches at y^k which do not contain p_2, $i = 2, 3, \cdots, m$. If x^kx^{k-1} and y^ky^{k-1} are the arcs from x^k and y^k to the $(k-1)$-st branch points, we define $T(x^kx^{k-1}) = y^ky^{k-1}$ and $T(x^{k-1}) = y^{k-1}$; also $T(b^i{}_{x,k-1}) = b^i{}_{y,k-1}$ where $b^i{}_{x,k-1}$ are the minor branches at x^{k-1} not containing x^k and $b^i{}_{y,k-1}$ the minor branches at y^{k-1} not containing y^k, $i = 2, 3, \cdots, j$. Continue in this manner until $x^j = y^j$ or until c is reached on both components. If b^1 and b^2 are the branches from x^j or c, as the case may be, which contain C_1 and C_2 respectively, for each point $x \epsilon b^2$, define $T(x) = T^{-1}(x)$; for each point $x \epsilon H - b^1 - b^2$, define $T(x) = x$. We thus have a homeomorphism T which sends C_1 into C_2, the component of $H - c$ which contains C_1 into the component of $H - c$ which contains C_2, and H into itself.

2.8. *Definition.* A set of points H is said to be *bi-cyclic element homogeneous* if, given any two true cyclic elements of H, there exists a homeomorphism T such that $T(H) = H$ and the true cyclic elements are sent into each other.

From the way the homeomorphism T was defined in 2.7, it is seen that if a set H satisfies the conditions of 2.7 it is bi-cyclic element homogeneous.

As was stated in the introduction, Fig. 3 is an example of a space M which is cyclic element homogeneous but it is not bi-cyclic element homogeneous. The true cyclic element C_1 may be sent into any other true cyclic element of the space by a homeomorphism which sends M into itself; but C_1 can not be sent into C_2 by a homeomorphism which sends C_2 into C_1 and M into itself.

It may be remarked that the finite case just treated could have been reduced to the consideration of an acyclic curve which was end point homogeneous. However, little if any advantage in simplicity seems to accrue from such a reduction.

3. *The case of infinitely many true cyclic elements.* Although no complete solution for this case of the problem has yet been obtained, we shall state here some results bearing on certain important phases of it.

3.1. If H is cyclic element homogeneous and contains infinitely many true cyclic elements which are grouped in clusters, where no cluster cuts H,

then there are a finite number of clusters with an infinite number of true cyclic elements in each cluster. Also under this hypothesis there exists a point c such that each cluster is the same number of branch points away from c and at each i-th branch point $(i = 1, 2, \cdots, k)$ from c there are the same number of branches. It can be shown that the necessary and sufficient condition for this case is similar to that of the finite case.

3.2. Let H satisfy the conditions: (1) it contains infinitely many true cyclic elements $\{C_j\}$, (2) each cut point is contained in exactly two true cyclic elements, (3) each component of $H - C_j$ has a different boundary point, (4) the set of cut points of H is totally disconnected, and (5) no point p is the limit of true cyclic elements in more than one component of $H - p$.

Under these restrictions, a necessary condition that H be cyclic element homogeneous is that when C_s and C_t are any two true cyclic elements of H, there exists a homeomorphism of C_s into C_t such that one particular cut point in C_s is sent into one particular cut point in C_t and the remaining cut points in C_s are sent into the remaining cut points in C_t.

Thus far it has not been shown that this condition is sufficient for H to be cyclic element homogeneous. But a sufficient condition is that when C_s and C_t are any two true cyclic elements of H, there exists a homeomorphism T of C_s into C_t such that two particular cut points in C_s are sent into two particular cut points in C_t and the remaining cut points in C_s are sent into the remaining cut points in C_t.

It is to be noted that the sufficient condition is stronger than the necessary condition.

THE JOHNS HOPKINS UNIVERSITY.

A NAVIGATION PROBLEM IN THE CALCULUS OF VARIATIONS.*

By E. J. McShane.

The problem which I shall consider in this note is closely related to the Zermelo Navigation problem.[1] Let us suppose that the velocities relative to the air which can be attained by an airship consist of all the vectors r lying in a convex body $K(x, t)$, depending on the position $x = (x^1, x^2, x^3)$ and the time t. The air is supposed to be in motion, its velocity being a continuous vector function $u(x, t)$. Given two points x_0, x_1, the problem is to find a path from x_0 to x_1 which can be traversed by the ship in the least possible time. If $K(x, t)$ is the sphere[2] $|r| \leqq k$, where k is a constant, and if we add the further requirement that the speed relative to the air shall almost always be exactly k, this becomes the Zermelo navigation problem. Our replacement of the sphere $|r| \leqq k$ by the convex body K is suggested by the fact that an airplane can travel faster down than up.

In the present paper I first prove under weak hypotheses the existence of a solution of the problem proposed above. I then consider the problem modified so as to be a generalization of the Zermelo problem (i. e. the ship's velocity is required to be almost always as great as possible), and under stronger hypotheses I prove that this problem also is solvable.

1. Throughout the following pages we shall use the following definitions and assumptions:

A is a bounded closed point set (atmosphere) in (x^1, x^2, x^3)-space.

Δ is a bounded closed set of real numbers t.

$K(x, t)$ is a bounded closed convex point set (or set of vectors) in three-dimensional space, defined and continuous[3] for $x \varepsilon A$ and $-\infty < t < \infty$.

$u(x, t)$ is a vector function $(u^1(x, t), u^2(x, t), u^3(x, t))$ defined and continuous for $x \varepsilon A$ and all t.

* Received December 3, 1936.

[1] For a detailed study of this problem, as well as a bibliography of previous papers, the reader is referred to a memoir by B. Manià, shortly to appear in *Mathematische Annalen*.

[2] $|r|$ is the length $[\Sigma(r^i)^2]^{1/2}$ of the vector r.

[3] For any convex set K, let K_ϵ be the set of all points having distance $\leqq \epsilon$ from K. A convex set $K(\theta)$ is a continuous function of θ at θ_0 if there is a neighborhood U of θ_0 such that for every θ in U the inclusions $K(\theta_0) < K_\epsilon(\theta)$ and $K(\theta) < K_\epsilon(\theta_0)$ hold.

$V(x, t)$ is the set of all vectors of the form $u(x, t) + r$ with $r \varepsilon K(x, t)$. That is, V is the translation of K by the vector u, and so it satisfies all the conditions imposed above on K.

Let the path of the ship be given in the form $C: x = x(t)$, $a \leqq t \leqq a + T$, where the parameter t is the time. The mean velocity of the ship between times t_1 and t_2 is $(x(t_2) - x(t_1))/(t_2 - t_1)$, and we shall assume that this is bounded. The velocity at time t is $x'(t)$, if $x'(t)$ exists. The velocity relative to the air is then $x'(t) - u(x(t), t)$, and this must be with the class $K(x(t), t)$ of velocities attainable at place $x(t)$ and time t. That is, by the definition of $V(x, t)$ we must have $x'(t) \varepsilon V(x(t), t)$. Combining this with the previous requirements on the paths $x(t)$ to be considered, we are led to the definition:

The curve C: $x = x(t)$, $a \leqq t \leqq a + T$ is *admissible* (or, more fully, an admissible curve traversable in the interval $[a, a + T]$) if

(1. 1a) the functions $x(t)$ are Lipschitzian;
(1. 1b) $x(t) \varepsilon A$ for $a \leqq t \leqq a + T$;
(1. 1c) $a \varepsilon \Delta$;
(1. 1d) $x(a) = x_0$, $x(a + T) = x_1$;
(1. 1e) $x'(t) \varepsilon V(x(t), t)$ wherever $x'(t)$ is defined.

It is convenient for the proof to have also the notation of a *weakly admissible* curve. A curve C is *weakly admissible* if it satisfies (1. 1a, b, c, d) and satisfies (1. 1e) if we replace the words "wherever $x'(t)$ is defined" by "almost everywhere."

We obtain an obviously equivalent definition by changing parameters from t to $\tau = (t - a)/T$; a curve C: $x = x(\tau)$, $0 \leqq \tau \leqq 1$ is an admissible curve traversable in the interval $[a, a + T]$ if

(1. 2a) the functions $x(\tau)$ are Lipschitzian;
(1. 2b) $x(\tau) \varepsilon A$ for $0 \leqq \tau \leqq 1$;
(1. 2c) $a \varepsilon \Delta$;
(1. 2d) $x(0) = x_0$, $x(1) = x_1$;
(1. 2e) $x'(\tau)/T \varepsilon V(x(\tau), a + T\tau)$ wherever $x'(\tau)$ is defined.

Likewise, if in (1. 2) we replace the words "wherever $x'(\tau)$ is defined" by "for almost all τ" we obtain a definition of a weakly admissible curve.

If a function $F(t)$ is defined and summable over a set $E - N$, where E is measurable and N has measure 0, we shall define

$$\int_E F(t)\,dt = \int_{E-N} F(t)\,dt.$$

This allows us to write, for example,

$$\int_a^b x'(t)\,dt = x(b) - x(a)$$

if $x(t)$ is absolutely continuous on $[a, b]$.

2. We now can state:

THEOREM I. *Under the hypotheses of § 1, if there exists an admissible curve, then there is an admissible curve C traversable in a time interval $[a, a+T]$ for which the time of traversal T is the least possible.*

Let T_0 be the greatest lower bound of the times of traversal of all admissible curves and let W be the lower bound of the times of traversal of all weakly admissible curves. Since the latter class contains the former, $W \leqq T_0$.

We now choose a sequence of weakly admissible curves C_n: $x = x_n(\tau)$, $0 \leqq \tau \leqq 1$, traversable in the respective intervals $[a_n, a_n + T_n]$, for which $T_n \to W$. From these we can select a subsequence such that a_n tends to a definite limit a_0; we suppose that $\{C_n\}$ is already such a sequence. Since Δ is closed, $a_0 \varepsilon \Delta$. All the time intervals $[a_n, a_n + T_n]$ lie in a bounded closed time interval, and all x lie in the bounded closed set A, and $V(x, t)$ is continuous; so for all such (x, t) the body $V(x, t)$ lies in a sphere about the origin of finite radius M. Then $| x'_n(\tau) |/T_n \leqq M$, so $| x'_n(\tau) |$ is uniformly bounded. Hence by Hilbert's theorem we can select a subsequence of the $x_n(\tau)$ which converges uniformly to a Lipschitzian limit function $x_0(\tau)$. We wish to prove that $x_0(\tau)$ is the curve sought. The proof will be given in a lemma.

LEMMA 2.1. *If the curves C_n: $x = x_n(\tau)$, $0 \leqq \tau \leqq 1$ are weakly admissible curves traversable in the intervals $[a_n, a_n + T_n]$, and*

$$(2.3) \qquad a_n \to a_0, \quad T_n \to U, \quad x_n(\tau) \to x_0(\tau) \text{ uniformly in } \tau,$$

then C_0: $x = x_0(\tau)$ is an admissible curve traversable in the time interval $(a, a + U)$.

Condition (1.2a) clearly holds, for $x_0(t)$ has bounded derivatives as we saw just above. (1.2b) holds by the closure of A, and (1.2c) by the closure of Δ. (1.2d) is evident, for $x_0 = x_n(0) \to x_0(0)$ and $x_1 = x_n(1) \to x_0(1)$. We now turn to the proof of (1.2e).

Suppose that τ_0 is any number in $[0, 1]$ such that $x'_0(\tau)$ exists. If ϵ is a positive number, we define V_ϵ to be the set of all points whose distance from

$V(x_0(\tau_0),\ a_0 + U\tau_0)$ is $\leqq \epsilon$. By (2.3) and the continuity of $x_0(\tau)$ and $V(x, t)$, we find that

(2.4) *there exists a* $\delta > 0$ *and an integer* n_0 *such that if* $|\tau - \tau_0| < \delta$ *and* $n > n_0$, *then* $V(x_n(\tau), a_n + T_n\tau) < V_\epsilon$. Therefore, for all $n > n_0$ and almost all τ such that $|\tau - \tau_0| < \delta$,

$$x'_n(\tau)\ \varepsilon\ V_\epsilon. \tag{2.5}$$

Now suppose $0 < |h| < \delta$. Since V_ϵ is closed and convex, by Jensen's inequality [4]

$$(x_n(\tau_0 + h) - x_n(\tau_0))/h = (1/h) \int_{\tau_0}^{\tau_0+h} x'_n(\tau)\,d\tau\ \varepsilon\ V_\epsilon.$$

Let h be fixed and let $n \to \infty$. By the closure of V_ϵ,

$$[x_0(\tau_0 + h) - x_0(\tau_0)]/h = \lim_{n\to\infty} [x_n(\tau_0 + h) - x_n(\tau_0)]/h\ \varepsilon\ V_\epsilon.$$

Now let $h \to 0$. Again by the closure of V_ϵ, $x'_0(\tau_0)\ \varepsilon\ V_\epsilon$. But here the point $x'_0(\tau_0)$ does not depend on ϵ, and it can only belong to V_ϵ for every $\epsilon > 0$ if it belongs to $V_0 \equiv V(x_0(\tau_0), a_0 + U\tau_0)$. Hence (1.2e) is satisfied and Lemma 2.1 is established.[5]

Returning to the proof of Theorem I, by Lemma 2.1, the curve C_0 is an admissible curve traversable in time $W \leqq T_0$. But the time of traversal of any admissible curve is $\geqq T_0$. Hence the time of traversal of $C_0 = W = T_0$. This completes the proof of Theorem I.

Remark. We could alter the problem by assuming that $u(x, t)$ and $K(x, t)$ are defined only for t in a closed interval $t_0 \leqq t \leqq t_1$; nothing in the preceding demonstration would be altered.

From Lemma 2.1 we draw another conclusion:

LEMMA 2.2. *Every weakly admissible curve is admissible.*

For let $C: x = x(\tau),\ 0 \leqq \tau \leqq 1$ be a weakly admissible curve traversable in the interval $(a, a + T)$. In Lemma 2.1 we take $x_n(\tau) = x_0(\tau) = x(\tau)$, $T_n = T = U$ for all n. Then the hypotheses of the lemma are satisfied, and the conclusion informs us that the limit curve (which is C itself) is an admissible curve traversable in the interval $[a, a + T]$.

---

[4] Cf., for example, E. J. McShane, "On Jensen's inequality," *Bulletin of the American Mathematical Society*, vol. 40 (1937).

[5] We have proved a little more than we stated. If τ_0 is arbitrary, the above proof shows that every vector which is the limit of a sequence $[x_0(\tau_0 + h_m) - x_0(\tau_0)]/h_m$ as $h_m \to 0$ must belong to $V(x(\tau_0), a + T\tau_0)$, even though $x'_0(\tau_0)$ may not exist.

3. For the next theorem we shall add the hypothesis that the ship's engines are powerful enough so that at any time and place it can proceed with speed $\geqq \delta$ (δ a positive number) in any desired direction. Analytically, this means that the set of velocities which the ship can attain, namely

$$V(x,t) = u(x,t) + K(x,t),$$

shall contain all velocities v for which $|v| \leqq \delta$; that is, the sphere $|v| \leqq \delta$ is contained in $V(x,t)$ for all x in A and all t. If this is the case, for each direction (unit vector) d and each (x,t) there is just one number $\rho = \rho(d,x,t)$ such that ρd is on the boundary of $V(x,t)$; and $\rho \geqq \delta$ for all d, all x in A and all t. It is quite easy to see that if $\rho(d,x,t)$ is continuous, then the body $V(x,t)$ is a continuous function of (x,t). It is somewhat less easy to see that the converse is true;[6] so in order to save some space we shall hereafter replace the assumption that $V(x,t)$ is continuous by the assumption (only apparently stronger) that $\rho(d,x,t)$ is continuous. Then

THEOREM II. *Let the boundary of the body $V(x,t)$ be given in polar coördinates by the equation $\rho = \rho(d,x,t)$, where $\rho(d,x,t) \geqq \delta > 0$ and ρ is a continuous function of all its arguments. Then, if there exists an admissible curve, there is an admissible curve $C: x = X(t)$, $a \leqq t \leqq a+T$ such that the time T of traversal is the least possible, and such moreover that*

$$0 < |X'(t)| = \rho(X'(t)/|X'(t)|, x(t), t)$$

for almost all t.

By Theorem I, there is a curve $C: x = x(t)$, $a \leqq t \leqq a+T$ which can be traversed in time T which is the least possible time of traversal of any admissible curve. Let $x = \xi(s)$, $0 \leqq s \leqq L$ be the representation of C with arc length as parameter. To each t in $(a, a+T)$ there corresponds a value $s_0(t)$ of the parameter s, and $s'_0(t) = |x'(t)|$ for almost all t. This function is monotonic increasing and absolutely continuous. It may not have a single valued inverse, but we define $t_0(s)$ to be the least t such that $s_0(t) = s$. Then $t_0(s)$ is defined and single valued (possibly discontinuous) for $0 \leqq s \leqq L$, and $s_0(t_0(s)) \equiv s$. Also

$$t_0(0) = a, \qquad t_0(L) = a+T. \tag{3.1}$$

Since $x'(t) \in V(x(t),t)$, we see that $|x'(t)|$ is bounded, say $\leqq M$. Then for any times t_1, t_2 we find

$$|s_0(t_1) - s_0(t_2)| = \left| \int_{t_0}^{t_1} s'_0(t)\,dt \right| \leqq M\,|t_0 - t_1|,$$

[6] This follows, for example, from pages 14 and 37 of Bonnesen and Fenchel, *Theorie der konvexen Körper*.

whence for all values $s_1, s_2 > s_1$ of s

$$(3.2) \qquad t_0(s_2) - t_0(s_1) \geqq (s_2 - s_1)/M.$$

The derivative $t'_0(s)$ exists and is finite for almost all s. Inequality (3.2) shows $t'_0 \geqq M^{-1}$, so

(3.3) *for almost all s the derivatives $\xi'(s)$, $t'_0(s)$, and $s'_0(t_0(s))$ exist and are finite, and $|\xi'(s)| = 1$, and $s'_0(t)t'_0(s) = 1$.*

From this and the identity $x(t) = \xi(s_0(t))$ it follows that for almost all s the derivative $x'(t)$ exists and

$$(3.4) \qquad x'(t) = \xi'(s_0)s'_0(t).$$

Let S be the set of all s (of measure L) on which (3.3) and (3.4) hold. We recall that for almost all t (and, a fortiori, almost all s) $x'(t) \in V(x(t), t)$, so almost everywhere in S

$$(3.5) \quad 0 < s'_0(t_0(s)) = |x'(t_0(s))| \leqq \rho(x'(t_0(s))/|x'(t_0(s))|, x(t_0(s)), t_0(s)) \\ = \rho(\xi'(s), \xi(s), t_0(s)).$$

This implies that almost everywhere in S

$$(3.6) \qquad t'_0(s) = [s'_0(t_0(s))]^{-1} \geqq [\rho(\xi'(s), \xi(s), t_0(s))]^{-1}.$$

The function $\rho(\xi'(s), \xi(s), t)$ is defined for almost all s, is measurable in s for fixed t (being a continuous function of the measurable functions $\xi'(s)$, $\xi(s)$) and is continuous in t for fixed s. Moreover $\rho \geqq \delta$. Hence for every $\epsilon > 0$ the function $[\rho(\xi'(s), \xi(s), t) + \epsilon]^{-1}$ is measurable in s for fixed t, continuous in t for fixed s, and bounded. Therefore the equation

$$(3.7) \qquad t_\epsilon(s) = a - \epsilon + \int_0^s [\rho(\xi'(s), \xi(s), t_\epsilon(s)) + \epsilon]^{-1} ds$$

has an absolutely continuous solution $t_\epsilon(s)$ on $(0, L)$. We now prove

(3.8) *If $0 \leqq \beta < \epsilon$, then $t_\beta(s) > t_\epsilon(s)$ $(0 \leqq s \leqq L)$.*

The graphs of $t = t_\beta(s)$ and $t = t_\epsilon(s)$ are continuous curves. The latter is obvious, for $t_\epsilon(s)$ is continuous. So is the former if $\beta > 0$. If $\beta = 0$, the graph can still be considered as the continuous curve $s = s_0(t)$, $a \leqq t \leqq a + T$. It is clear that (3.8) holds at $s = 0$, for

$$t_\beta(0) - t_\epsilon(0) = (a - \beta) - (a - \epsilon) > 0.$$

If (3.8) is not always true, the graph of $t = t_\epsilon(s)$ lies somewhere above the graph of $t = t_\beta(s)$, and so they must have a first intersection point (σ, τ), $s_\epsilon(\tau) = s_\beta(\tau) = \sigma$. (Here $s_\epsilon(\tau)$ is the inverse of $t_\epsilon(s)$.) Thus for $s < \sigma$

$$t_\beta(s) > t_\epsilon(s). \tag{3.9}$$

By the uniform continuity of ρ, there is an $h > 0$ such that

$$\rho(\xi'(s), \xi(s), t_1) < \rho(\xi'(s), \xi(s), t_2) + \epsilon - \beta \text{ if } | t_1 - t_2 | < h. \tag{3.10}$$

By the continuity of $t_\epsilon(s)$, there is a $k > 0$ such that

$$0 \leqq t_\epsilon(\sigma) - t_\epsilon(s) < h \text{ if } \sigma - k \leqq s \leqq \sigma. \tag{3.11}$$

So for $\sigma - k \leqq s < \sigma$ we have, by (3.9) and (3.11),

$$\tau - h = t_\epsilon(\sigma) - h \leqq t_\epsilon(s) < t_\beta(s) \leqq \tau,$$

and therefore, by (3.10), for almost all s in $(\sigma - k, \sigma)$

$$\rho(\xi'(s), \xi(s), t_\beta(s)) + \beta < \rho(\xi'(s), \xi(s), t_\epsilon(s)) + \epsilon. \tag{3.12}$$

By (3.12) and (3.7) (and (3.6) if $\beta = 0$)

$$\tau'_\beta(s) > \tau'_\epsilon(s) \tag{3.13}$$

for almost all s in $(\sigma - k, \sigma)$. Now τ_β is monotonic increasing and τ_ϵ is absolutely continuous, so,[7] by (3.13)

$$\begin{aligned}\tau_\beta(\sigma) - \tau_\epsilon(\sigma) &\geqq \tau_\beta(\sigma - k) + \int_{\sigma-k}^{\sigma} \tau'_\beta(s)\,ds - \tau_\epsilon(\sigma - k) - \int_{\sigma-k}^{\sigma} \tau'_\epsilon(s)\,ds \\ &> \tau_\beta(\sigma - k) - \tau_\epsilon(\sigma - k) > 0.\end{aligned}$$

Since τ_ϵ is continuous and τ_β is monotonic increasing, we can find an interval $\sigma \leqq s \leqq \sigma + l$ on which

$$\tau_\beta(s) - \tau_\epsilon(s) \geqq \tau_\beta(\sigma) - \tau_\epsilon(s) > 0.$$

So the graph of τ_β lies above the graph of τ_ϵ for $0 \leqq s \leqq \sigma + l$, and (σ, τ) cannot be an intersection point. This contradiction establishes statement (3.8).

Now let ϵ tend to 0 through a monotonic decreasing sequence of values. By (3.8) the successive functional values $t_\epsilon(s)$ increase for each s, so there exists a limit function:

$$\lim_{\epsilon \to 0} t_\epsilon(s) = \tau(s), \qquad 0 \leqq s \leqq L. \tag{3.14}$$

---

[7] Hobson, *Theory of Functions of a Real Variable*, vol. I, p. 590.

Also by (3.8), $t_\epsilon(s) \leqq t_0(s)$ for each $\epsilon > 0$, so in the limit

$$\tau(s) \leqq t_0(s), \qquad 0 \leqq s \leqq L. \tag{3.15}$$

In equation (3.7), the integrand on the right is uniformly bounded and tends almost everywhere to $[\rho(\xi'(s), \xi(s), \tau(s))]^{-1}$, while on the left $t_\epsilon \to \tau$; so

$$\tau(s) = a + \int_0^s [\rho(\xi'(s), \xi(s), \tau(s))]^{-1} ds. \tag{3.16}$$

The function $\tau(s)$ has almost everywhere a positive derivative, by (3.16), so it has an absolutely continuous inverse $s = s(\tau)$, $a \leqq \tau \leqq \tau(L)$. Defining $X(\tau) = \xi(s(\tau))$, we find that for almost all τ

$$X'(\tau) = \xi'(s)s'(\tau), \qquad s'(\tau) > 0,$$

so $X'(\tau)/|X'(\tau)| = \xi'(s)$ for almost all τ. Moreover, for almost all τ

$$\begin{aligned} X'(\tau) &= \xi'(s)[\tau'(s)]^{-1} = \xi'(s)\rho(\xi'(s), \xi(s), \tau(s)) \\ &= \xi'(s(\tau))\rho(X'(\tau)/|X'(\tau)|, X(\tau), \tau), \end{aligned}$$

whence

$$|X'(\tau)| = \rho(X'(\tau)/|X'(\tau)|, X(\tau), \tau)$$

for almost all τ. That is, condition (2.1e) holds with the words "wherever $X'(\tau)$ is defined" replaced by "for almost all τ." The others of conditions (2.1) are trivially easily verified, so $x = X(\tau)$ is weakly admissible, and by Lemma 2.2 it is admissible. All that remains to prove is that the time of traversal $\tau(L) - a$ has the least possible value T. Clearly $\tau(L) - a \geqq T$. By (3.15) and (3.1), $\tau(L) \leqq t_0(L) = a + T$, completing the proof.

Remark. If we change the problem by assuming that $u(x, t)$ and $K(x, t)$ are defined only on a time interval $t_0 \leqq t \leqq t_1$, one point of the preceding demonstration needs change. In defining t_ϵ, times $t < a$ entered. If however we define $V(x, t) = V(x, t_0)$ for $t < t_0$, the proof can be carried out as above, and the final results will involve only times t in $[t_0, t_1]$.

UNIVERSITY OF VIRGINIA.

THE TOPOLOGICAL DISCRIMINANT GROUP OF A RIEMANN SURFACE OF GENUS p.

By Oscar Zariski.

1. Introduction. The symmetric n-th product K^n of a complex K carries a subset D whose points represent n-tuples on K with two or more coincident points. We call D the *discriminant variety* of K^n and we refer to the fundamental group of the residual space $K^n - D$ as the *topological discriminant group* (*of degree* n) of the given complex K. In part I we determine this group, $G_{n,p}$, when K is a Riemann surface R of genus p. We were led to examine this group by the following considerations. The variety R^n is the space of all n-tuples of points of an algebraic curve f of genus p. As such, R^n carries—for a sufficiently high value of n $(n \geqq p+2)$—a system, ∞^p, of linear $(n-p)$-spaces S_{n-p}, images of complete linear series g_n^{n-p} existent on f. If D_1 denotes the intersection of a general S_{n-p} of the system with the discriminant variety D of R^n, the fundamental group of the residual space $S_{n-p} - D_1$ can be shown to be an invariant subgroup $H_{n,p}$ of $G_{n,p}$, and the quotient group $G_{n,p}/H_{n,p}$ is simply isomorphic to the homology group of R. If $G_{n,p}$ is known, $H_{n,p}$ can be determined on the basis of well-known principles laid down by Reidemeister.[1] By a theorem which we have proved elsewhere (Zariski [5]) the fundamental group of $S_{n-p} - D_1$ coincides with the fundamental group of the residual space of a general plane section C of D_1. It is not difficult to see that C is the plane dual of a general plane curve of order n and genus p, so that C is of order $2n+2p-2$ with $3(n+2p-2)$ cusps and $2(n-2)(n-3)+2p(2n+p-7)$ nodes. The knowledge of the fundamental group of C, of interest in itself, makes it also possible to determine the fundamental group of any plane curve admitting C as a limiting case. In this connection we may point out that the class of curves thus obtained is not negligible, since, at present, duality constructions and limiting processes are, with a few exceptions, the only means of arriving at effectively existent curves with nodes and cusps.

In Parts II and III we carry out in detail the above outlined considerations in the case $p=1$. The somewhat elaborate group-theoretic apparatus of Part II is inherent to the reduction of the infinite set of generators and generating relations of $H_{n,1}$ (an invariant subgroup of $G_{n,1}$ of infinite index) to a finite set of generators and generating relations. The existence of such a

finite set is, *a priori*, implied by the algebro-geometric interpretation of $H_{n,1}$ given in Part III.

An interesting special case, examined in Part III, is given by the dual of a plane cubic—a sextic with 9 cusps. It is then found that the 9 generating relations at the cusps enjoy properties which are in striking analogy with the well-known alignment properties of the configuration of the 9 flexes of a plane cubic.

2. Let R be a Riemann surface of genus p and let R^n be the symmetric n-th product of R, i. e. the space (of n complex dimensions) of all unordered n-tuples of points of R, topologized in an obvious manner. We have shown elsewhere (Zariski,[3] p. 1), that R^n is a manifold. We denote by D the subvariety of R^n whose points correspond to n-tuples of points of R in which two or more points coincide. This variety D—which can legitimately be designated as the *discriminant variety of* R^n—is of $n-1$ complex dimensions. The purpose of this and of the next two sections is the determination of the fundamental group of the residual space $R^n - D$. We denote this group by $G_{n,p}$.

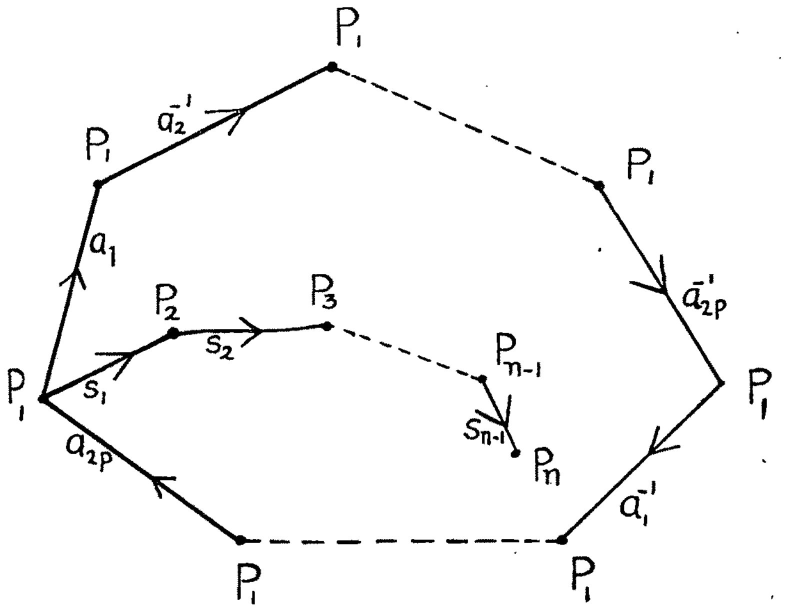
The group $G_{n,0}$ is known (see Zariski,[4] p. 612). As in the just quoted paper, we interpret also here the group $G_{n,p}$ as the *group of motion classes of* n *points of* R. The motions considered are those which carry a fixed initial set of n distinct points $P_1, P_2, \cdots, P_n$ of R into its initial position (allowing for a permutation of the points P_i) and in the course of which the variable set consists always of distinct points. Two motions belong to the same class if they can be deformed into each other through a continuous chain of motions of the same nature.

We fix on R a set of retrosections $a_1, a_2, \cdots, a_{2p}$ on a common point P_1, belonging to our initial n-tuple of points. We choose our retrosections in such a manner that when R is cut open along them, the resulting 2-cell E_2 is bounded by the closed polygon

$$a_1 a_2^{-1} a_3 \cdots a_{2p}^{-1} a_1^{-1} a_2 a_3^{-1} \cdots a_{2p}.$$

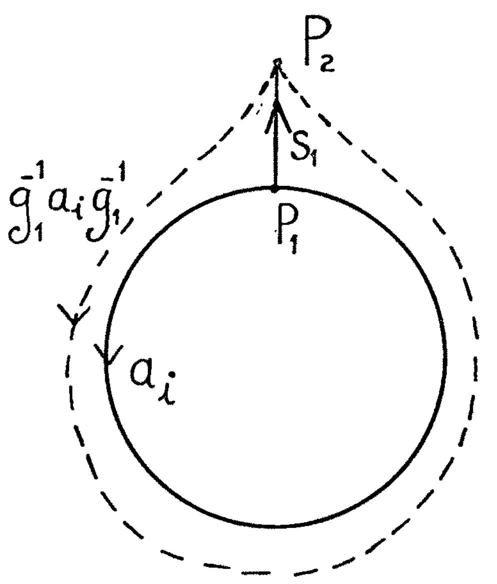
We assume the points $P_2, \cdots, P_n$ in the interior of E_2 and we join the points $P_1, P_2, \cdots, P_n$ by a set of simple oriented arcs $s_1, s_2, \cdots, s_{n-1}$ (see figure 1). The indicated orientations of the retrosections a_i and of the arcs s_j are such that at P_1 the positive sense on each retrosection a_i points from the right-hand edge of the arc s_1 toward its left-hand edge (see figure 2).

We denote by g_i the motion in which P_i is carried into P_{i+1} along the left-hand edge of s_i and P_{i+1} is carried into P_i along the right-hand edge of s_i, while the remaining points P_k are fixed in their initial positions. In the case



P_1
P_1
$\bar{a}_2^{\,1}$
P_1
a_1
$\bar{a}_{2p}^{\,1}$
P_2
P_3
s_2
P_{n-1}
P_1
s_1
P_1
s_{n-1}
a_{2p}
P_n
a_1^{-1}
P_1
P_1

Fig. 1.



P_2
S_1
$\bar{g}_1^{1} a_i \bar{g}_1^{1}$
P_1
a_i

Fig. 2.

$p=0$ the elements $g_1, g_2, \cdots, g_{n-1}$ are generators of $G_{n,0}$ (see Zariski,[4] p. 610). It follows that any motion in the course of which the points of the variable set do not cross the boundary of E_2 can be expressed as a product of the g_i's. Let us consider the motion in which P_1 describes the oriented retrosection a_i, while the remaining points P_k are fixed $(k>1)$. We shall denote this motion by the same letter a_i. It is obvious that any crossing of the boundary of E_2 introduces factors $a_i^{\pm 1}$, hence the *elements* $g_1, \cdots, g_{n-1}, a_1, \cdots, a_{2p}$ *are the generators of* $G_{n,p}$.

3. The generating relations of $G_{n,p}$. The relations

$$(\alpha) \qquad g_ig_j = g_jg_i, \qquad |i-j| \neq 1;$$

$$(\beta) \qquad g_ig_{i+1}g_i = g_{i+1}g_ig_{i+1}, \qquad (i=1,2,\cdots,n-2);$$

established in our quoted paper (Zariski,[4] p. 612) remain valid also in the present case. The relation (6) of the quoted paper now has to be replaced by the following:

$$(\gamma) \qquad g_1 \cdots g_{n-2}g^2_{n-1}g_{n-2} \cdots g_1 = a_1a_2^{-1}a_3 \cdots a_{2p}^{-1}a_1^{-1}a_2 \cdots a_{2p},$$

since the left-hand member represents a motion which, as is easily seen, can be deformed into one in which the points $P_2, \cdots, P_n$ are fixed, while P_1 describes a closed path surrounding the set $\{P_2, \cdots, P_n\}$. This closed curve can be deformed into the boundary of the cell E_2. Other generating relations, involving the g's and the elements a_i, are obtained as follows:

In the first place it is clear that each a_i is permutable with each of the elements $g_2, \cdots, g_{n-1}$, since the corresponding paths do not intersect. Hence

$$(\delta) \qquad g_ka_i = a_ig_k, \qquad (i=1,2,\cdots,2p;\ k=2,\cdots,n-1).$$

The motion $g_1^{-1}a_ig_1^{-1}$ can be deformed into a motion in which the points $P_1, P_3, \cdots, P_n$ are fixed, while P_2 describes a retrosection homologous to a_i and not meeting a_i (see figure 2, the path of P_2 is indicated by the punctuated curve). Hence a_i and $g_1^{-1}a_ig_1^{-1}$ are permutable, whence the relation

$$(\epsilon) \qquad (g_1^{-1}a_i)^2 = (a_ig_1^{-1})^2.$$

We now introduce the following elements:

$$(\mu) \qquad a'_i = g_1^{-1}a_ig_1.$$

It is clear that the motion a'_i is equivalent to a motion in which $P_1, P_3, \cdots, P_n$ are fixed, while P_2 describes a retrosection a'_i homologous to a_i, and that a'_i

and a_j intersect in one point only, provided $j > i$ (see figure 3, illustrating the behavior of a'_2). If we then consider the direct product of a'_i and a_j, regarded as 1-spheres, we have a torus T, on which the motions a_j and a'_i, regarded as motions of the point pair P_1P_2 (the remaining points $P_3, \cdots, P_n$ being fixed), are represented by two retrosections α and α' respectively. The common point $\bar{P}$ of α' and α corresponds to the initial point-pair (P_1, P_2). Let Q be the point at which a'_i and a_j intersect, and let $\bar{Q}$ be the point of the torus which corresponds to the point pair (Q, Q). A closed path on $T - \bar{Q}$ starting from and returning to $\bar{P}$ represents a motion of a variable pair of

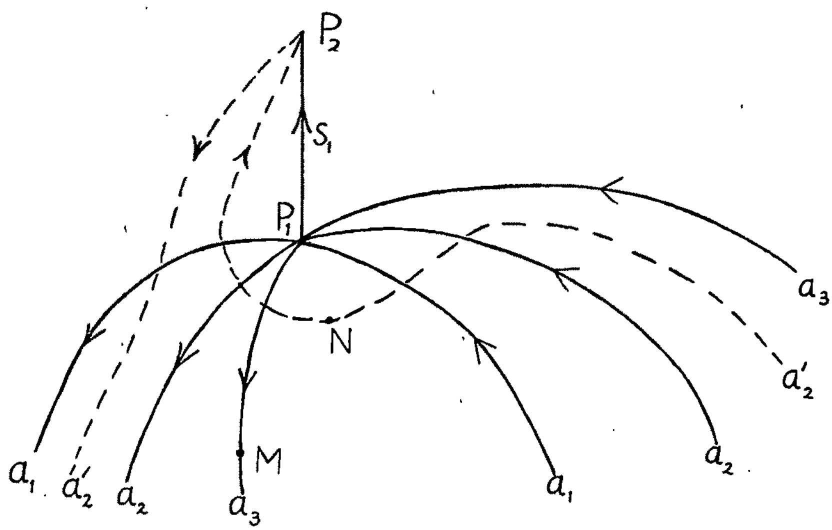


FIG. 3.

distinct points starting from and returning to the initial point pair P_1, P_2. A deformation of this path on $T - \bar{Q}$ corresponds to an allowable deformation of the corresponding motion on R. Since on $T - \bar{Q}$ we have $\alpha\alpha'^{-1}\alpha^{-1}\alpha' = \bar{\gamma}$, where $\bar{\gamma}$ is a properly oriented loop issued from $\bar{P}$ and surrounding the point $\bar{Q}$, we have a corresponding relation $a_j a'^{-1}_i a_j^{-1} a'_i = \gamma$, where γ is the motion of the variable point pair on R which corresponds to the loop $\bar{\gamma}$ on T. To determine γ, we take as $\bar{\gamma}$ a quadrangle two of whose sides are on the retrosections α and α' and the other two are parallel to these retrosections (see Fig. 4). The corresponding motion γ has now the following description: (a) first the point P_1 describes the arc P_1M on a_j, P_2 is fixed; (b) then P_2 describes the arc P_2N on a'_i, while the second point is fixed at M; (c) a reversal of motion (a); (d) the reversal of motion (b) (see Fig. 3, where $j = 3$, $i = 2$).

By letting M approach P_1 on a_j and by accompanying this by a deformation of the path (b), we see immediately that the combined motion γ can be deformed into a motion in which P_1 is fixed and in which P_2 turns around P_1 in what on Fig. 3 would be the clockwise sense. This motion is visibly equivalent to the motion g_1^2. We have therefore the following generating relation:

$$(\nu) \qquad a_j a'^{-1}_i a_j^{-1} a'_i = g_1^2, \qquad j > i.$$

We prove in the next section that *the relations* (α), (β), (γ), (δ), (ϵ) *and* (ν)

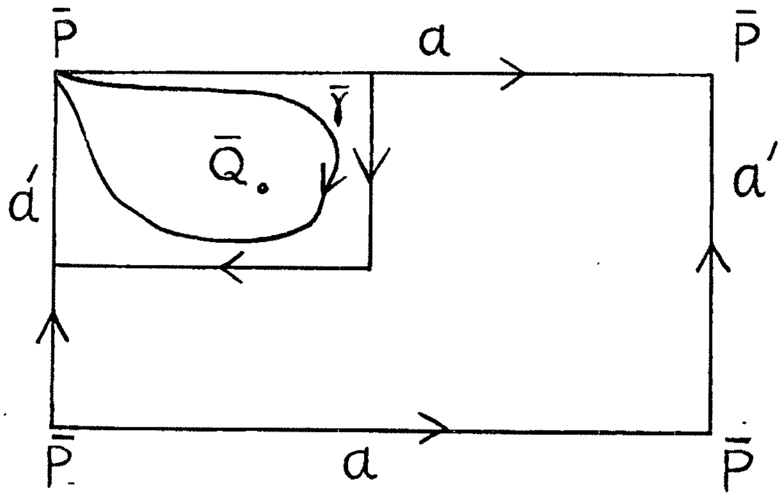


FIG. 4.

(where the elements a'_i are defined by (μ)) *constitute a complete set of generating relations of* $G_{n,p}$.

4. We denote the abstract group defined by the relations $(\alpha - \nu)$ of the preceding section by $\bar{G}_{n,p}$, and we use the notations $\equiv$ and $=$ to indicate equality of elements in $\bar{G}_{n,p}$ and $G_{n,p}$ respectively. We wish to prove that $\bar{G}_{n,p}$ coincides with $G_{n,p}$, or what is the same that $\alpha = \beta$ implies $\alpha \equiv \beta$, where α and and β are products of the generators a_i, g_k. The proof will be made in several steps.

(a) *If W is any product of the generators a_i, g_k, then $W \equiv g_h g_{h-1} \cdots g_1 W_1$, where $0 \leq h \leq n-1$ and where W_1 involves only the elements $g_2, \cdots, g_n, a_i, a'_i$ and the elements*

$$s_j = (g_{j-2} g_{j-3} \cdots g_1)^{-1} g^2_{j-1} (g_{j-2} g_{j-3} \cdots g_1), \qquad (j = 2, \cdots, n).$$

The proof is the same as the one given in Zariski,[4] pp. 612-613, except that it is also necessary to make use of the relations

$$a_i^{\pm 1} g_h \cdots g_1 = g_h \cdots g_2 a_i^{\pm 1} g_1 = g_h \cdots g_2 g_1 a'^{\pm 1}_i.$$

The elements s_j, denoted in Zariski,[4] p. 612, by a_j, represent motions in which the points $P_2, \cdots, P_n$ are at rest while P_1 describes a loop around the point P_j.

(b) *If W represents a motion in which P_1 returns to its initial position, then $W \equiv W_1$.*

This is a consequence of (a) since in $g_h \cdots g_1 W_1$ the point P_{h+1} is carried into P_1.

(c) *The subgroup Γ of $\bar{G}_{n,p}$ generated by the elements $s_2, \cdots, s_n, a_1, \cdots, a_{2p}$ is an invariant subgroup of the group generated in $\bar{G}_{n,p}$ by the above elements and by the elements $g_2, \cdots, g_{n-1}, a'_i$.*

That $g_k^{\pm 1} s_j g_k^{\mp 1} \subset \Gamma$ for $k = 2, \cdots, n-1$ has already been proved in Zariski,[4] p. 613. In view of (δ) it remains to prove that $a'^{\pm 1}_i \Gamma a'^{\mp 1}_i \subset \Gamma$.

The relation (ϵ) shows that a'_i is commutative with $s_2^{-1} a_i$ $(= g_1^{-2} a_i)$. Hence either one of the two relations

$$(1) \qquad a'^{\epsilon}_i s_2 a'^{-\epsilon}_i \subset \Gamma, \qquad a'^{\epsilon}_i a_i a'^{-\epsilon}_i \subset \Gamma, \qquad \epsilon = \pm 1$$

implies the other. Now (ϵ) also implies the relation $(a'_i)^{-1} a_i a'_i = s_2^{-1} a_i s_2$. Consequently both relations (1) hold true for $\epsilon = -1$. Transforming the relation $(a'_i)^{-1} a_i a'_i = s_2^{-1} a_i s_2$ by $(a'_i)^{-1}$ and taking into account the commutativity of the elements a'_i, $s_2^{-1} a_i$, we find that $a'_i s_2 a'^{-1}_i$ belongs to Γ. Hence the relations (1) hold true for $\epsilon = \pm 1$.

Since $g_1 s_j g_1^{-1}$, $j > 2$, involves only the elements $g_2, \cdots, g_{n-1}$, it follows by (δ) that $g_1 s_j g_1^{-1}$ is commutative with each a_i, and hence s_j is commutative with a'_i $(= g_1^{-1} a_i g_1)$, for $j > 2$.

It remains to prove that all the transforms $(a'_i)^{\epsilon} a_j (a'_i)^{-\epsilon}$, $\epsilon = \pm 1$, $i \neq j$, belong to Γ. For $\epsilon = -1$ and $i < j$ this follows directly from the relation (ν), and for $\epsilon = +1$ and $i < j$ this is proved by transforming the relation (ν) by $(a'_i)^{-1}$, since we have already proved that $a'_i g_1^2 a'^{-1}_i$ is in Γ. We now transform (ν) by g_1 and we obtain the relation $a'_j s_2^{-1} a_i s_2 a'^{-1}_j = s_2 a_i^{-1}$. Since $a'^{\epsilon}_j s_2 a'^{-\epsilon}_j \subset \Gamma$, for $\epsilon = \pm 1$, it follows immediately that $a'^{\epsilon}_j a_i a'^{-\epsilon}_j \subset \Gamma$, and this completes the proof of the invariance of the subgroup Γ.

(d) As a consequence of (b) and (c) we may now assert that if W represents a motion in which the point P_1 returns to its initial position, then we have already in $\bar{G}_{n,p}$ a relation of the form:

$$(2) \quad W \equiv W_1(g_2, \cdots, g_{n-1};\ a'_1, \cdots, a'_{2p}) \cdot W_2(s_2, \cdots, s_{n-1};\ a_1, \cdots, a_{2p}),$$

where W_1 and W_2 are products of the elements indicated in the parentheses.

(e) To prove that the groups $G_{n,p}$ and $\bar{G}_{n,p}$ are identical, we use an induction with respect to n, since for $n = 1$ the group $G_{n,p}$ is merely the fundamental group of the Riemann surface R, and in this case the relation (ν), where now the left-hand member reduces to 1, is the only generating relation for $G_{n,p}$. Hence $G_{1,p}$ coincides with $\bar{G}_{1,p}$. We shall then assume that $G_{n-1,p}$ and $\bar{G}_{n-1,p}$ are identical groups. For $G_{n-1,p}$ we take as initial sets of $n-1$ points the points $P_2, \cdots, P_n$, and as generators the elements $g_2, \cdots, g_{n-1}$, $a'_1, \cdots, a'_{2p}$. As elements analogous to $g_1^{-1}a_ig_1$ $(= a'_i)$ we take the elements $a''_i = g_2^{-1}a'_ig_2$.

Let W be an element of $G_{n,p}$, expressed as a product of the generators g_k, a_i, and let $W = 1$ be a true relation in $G_{n,p}$. Since in the motion W every point P_i returns to its initial position, the representation (2) of W holds true. Since in the motion W_1 the point P_1 is fixed, while in the motion W_2 the points $P_2, \cdots, P_n$ are fixed, it is clear that $W_1 = 1$ must be a true relation in $G_{n-1,p}$. By our induction, this relation must be a consequence of the relations $(\alpha - \nu)$ for the case $n - 1$. To rewrite these relations for the group $G_{n-1,p}$ we must replace $g_1, \cdots, g_n$ by $g_2, \cdots, g_n$ and the elements a_i by the elements a'_i. Let us letter these new relations by $\alpha', \beta', \cdots, \nu'$.

We assert that the relations (α'), (β'), (δ'), (ϵ'), (ν') *are group-theoretic consequences of the relations* (α), (β), (δ), (ϵ), (ν). The relations (α'), (β') are among the relations (α), (β). As for the relations (δ'), we observe that by (α), g_1 is commutative with g_k, $k \geq 3$, and hence the relations (δ') are obtained by transforming by g_1 those relations (δ) in which $k \geq 3$. Finally, the relations (ϵ'), (ν') are the transforms of (ϵ) and (ν) by g_2g_1. In fact, since g_2 and a_i are commutative, we have

$$(g_2g_1)^{-1}a_i(g_2g_1) = g_1^{-1}a_ig_1 = a'_i.$$

Taking into account the relation $g_1g_2g_1 = g_2g_1g_2$ we find

$$(g_2g_1)^{-1}a'_i(g_2g_1) = (g_1g_2)^{-1}a_i(g_1g_2) = g_2^{-1}a'_ig_2 = a''_i.$$

Moreover, we have $(g_2g_1)^{-1}g_1(g_2g_1) = g_2$, as a consequence of the relation $g_1g_2g_1 = g_2g_1g_2$.

We have left out the relation (γ'), i.e. the following:

$$H = a'_1a'^{-1}_2 \cdots a'^{-1}_{2p}a'^{-1}_1a'_2 \cdots a'_{2p}(g_2 \cdots g^2_{n-1} \cdots g_2)^{-1} = 1.$$

This is not a true relation in $\bar{G}_{n,p}$. In fact, if we transform (γ) by g_1, we find the following relation:

$$H \equiv g_1^2; \quad \text{i. e.,} \quad H \equiv s_2.$$

Having thus proved that all the generating relations of $G_{n-1,p}$, except the relation $H = 1$, are also true relations in $\bar{G}_{n,p}$, and recalling that $W_1 = 1$ holds in $G_{n-1,p}$ we deduce that W_1, as an element of the group $\bar{G}_{n,p}$, can be expressed as a product of transforms of H, i. e. of s_2, the transforming elements involving only the generators $g_2, \cdots, g_n, a'_i$ of $G_{n-1,p}$. Hence, by (c), we can write $W_1 \equiv W'_1$, where $W'_1 \subset \Gamma$. In view of (2), we conclude that *if a relation* $W = 1$ *holds true in* $G_{n,p}$, *then we have in* $\bar{G}_{n,p}$ $W \equiv F(a_1, \cdots, a_{2p}; s_2, \cdots, s_n)$, *where* F *is a product involving only the element* a_i, s_j.

The elements $a_1, \cdots, a_{2p}, s_2, \cdots, s_n$ are generators of the fundamental group G^* of the Riemann surface with $n-1$ holes at $P_2, \cdots, P_n$, i. e. of $R - P_2 - \cdots - P_n$. It can be proved, as in Zariski,[4] p. 614, that the relation $F = 1$, true in $G_{n,p}$, implies that F, considered as an element of G^*, belongs to the center of G^*. Since G^* is a free group, for $n \geqq 2$, it follows that *F is the identity in* G^*. The only generating relation of G^* is the following:

$$a_1 a_2^{-1} \cdots a_{2p}^{-1} a_1^{-1} a_2 \cdots a_{2p} = s_2 s_3 \cdots s_n.$$

If the s_j's are replaced by their expressions in terms of the g_k's, this relation coincides with the relation (γ). Hence $F = 1$ is also a true relation in $\bar{G}_{n,p}$, i. e. we have $F \equiv 1$. Consequently $W = 1$ implies $W \equiv 1$, q. e. d.

II. On an invariant subgroup of $G_{n,p}$ in the elliptic case.

5. In a motion which carries the initial n-tuple $P_1, \cdots, P_n$ back to its initial position, the paths described by the points $P_1, \cdots, P_n$ constitute together a closed curve, a singular 1-cycle σ_1. Those elements of $G_{n,p}$ for which this cycle σ_1 is ~ 0 on the Riemann surface R form an invariant subgroup $H_{n,p}$ of $G_{n,p}$, and the quotient group $G_{n,p}/H_{n,p}$ is the homology group of R. There is a general procedure, given by Reidemeister,[1] for determining the generators and the generating relations (finite or infinite in number) of an invariant subgroup of a discrete infinite group, whether the quotient group is finite or infinite. We shall now apply this general method to the invariant group $H_{n,p}$ of $G_{n,p}$ in the elliptic case ($p = 1$). It will be seen that $H_{n,1}$ admits a finite set of generators satisfying a finite set of relations, although the quotient group is in this case a free abelian group. This could also be foreseen from the geometric considerations of Part III of this paper, where it will be shown that the group $H_{n,1}$ is the fundamental group of the residual space of a certain elliptic plane curve.

The reduction of the set of generators and of generating relations of $H_{n,1}$, which we are about to undertake, can be extended to the group $H_{n,p}$, p arbitrary, in so far at least as the explicit determination of a finite set of generators is concerned. As for the generating relations, a similar reduction presents some difficulties.

From now on we shall denote the groups $G_{n,1}$ and $H_{n,1}$ by G_n and H_n respectively. G_n is generated by the elements $a_1, a_2, g_1, \cdots, g_{n-1}$. We rewrite the generating relations of G_n as follows:

$$(3)\quad \begin{cases} T_{ij} \;\;= g_i g_j g_i^{-1} g_j^{-1} = 1, & |i-j| \neq 1; \\ T_{i,i+1} = g_i g_{i+1} g_i g_{i+1}^{-1} g_i^{-1} g_{i+1}^{-1} = 1; & \\ T \;\;= g_1 \cdots g_{n-2} g^2_{n-1} g_{n-2} \cdots g_1 a_2^{-1} a_1 a_2 a_1^{-1} = 1; & \\ P_{ik} \;\;= a_i g_k a_i^{-1} g_k^{-1} = 1, & (k = 2, \cdots, n-1;\ i = 1, 2), \\ S_i \;\;= (a_i g_1^{-1})^2 (g_1^{-1} a_i)^{-2} = 1, & (i = 1, 2); \\ S_{12} \;\;= a_2 a'^{-1}_1 a_2^{-1} a'_1 g_1^{-2} = 1, \qquad a'_1 = g_1^{-1} a_1 g_1. & \end{cases}$$

Since a_1, a_2, considered as 1-cycles on R, are generators of the homology group, it follows that there is a $(1-1)$ correspondence between the elements $a_1^k a_2^l$ of G_n and the elements of the quotient group G_n/H_n. It is also clear that the elements of H_n are those and only those elements of G_n which become equal to 1 if the relations $g_i = 1$, $a_1^{-1} a_2 a_1 a_2^{-1} = 1$ are added to the generating relations of G_n, i. e. those power products of the generators a_1, a_2, g_k in which the sum of the exponents of each of the elements a_1, a_2 vanishes. It follows, by the quoted paper of Reidemeister, that the following elements are generators of H_n:

$$(4)\quad \begin{cases} \alpha_{ij} \;\;= (a_1^i a_2^j) a_1 (a_1^{i+1} a_2^j)^{-1}, & \\ \alpha'_{ij} \;\;= (a_1^i a_2^j) a_2 (a_1^i a_2^{j+1})^{-1}, & (i, j = 0, \pm 1, \pm 2, \cdots). \\ g_{ij} \;\;= (a_1^i a_2^j) g_1 (a_1^i a_2^j)^{-1}, & \\ g_{ij,k} = (a_1^i a_2^j) g_k (a_1^i a_2^j)^{-1}, & (k = 2, \cdots, n-1). \end{cases}$$

The elements α'_{ij} are all identically equal to 1, and hence we are left with the generators $\alpha_{ij}, g_{ij}, g_{ij,2}, \cdots, g_{ij,n-1}$.

Generating relations of H_n are obtained as follows. In the first place a *definite construction* is given by means of which any power product π of the generators of G_n can be expressed in the form $\pi_1 a_1^i a_2^j$, where π_1 is a power product of the generators of H_n. Here the exponents i, j are uniquely determined by π, since π and $a_1^i a_2^j$ correspond to one and the same element of the quotient group G_n/H_n. This construction is as follows. Let $\pi = \pi' \lambda^{\pm 1}$, where λ is a generator of G_n and where π' contains less factors than π, and let us assume that we have already expressed π' in the form $\pi'_1 a_1^i a_2^j$. Among the generators (4) of H_n there is an element λ_{ij} of the form $(a_1^i a_2^j) \lambda (a_1^{i'} a_2^{j'})^{-1}$

and there is also an element $\bar{\lambda}_{ij}$ of the form $(a_1{}^{i''}a_2{}^{j''})\lambda(a_1{}^{i}a_2{}^{j})^{-1}$. Then, if $\pi = \pi'\lambda$, we write $\pi = \pi'_1\lambda_{ij}a_1{}^{i'}a_2{}^{j'}$ and if $\pi = \pi'\lambda^{-1}$ we write $\pi = \pi'_1\bar{\lambda}_{ij}^{-1}a_1{}^{i''}a_2{}^{j''}$.

Using this construction, we obtain all the generating relations of H_n as follows: (a) we express the power products which occur in the relations (3) by means of the generators of H_n and we put equal to 1 the resulting expressions; (b) we apply the same procedure to the transforms of the relations (3) by any of the elements $a_1{}^{i}a_2{}^{j}$; (c) we finally apply our construction to the elements $a_1{}^{i}a_2{}^{j}$, getting $a_1{}^{i}a_2{}^{j} = \pi_{ij}a_1{}^{i}a_2{}^{j}$, where π_{ij} is a power product of the generators of H_n, and we put $\pi_{ij} = 1$.*

It is immediately seen that the relations $\pi_{ij} = 1$ give only the following trivial relations: †

$$\alpha'_{ij} = 1, \text{ for all } i \text{ and } j;$$
$$\alpha_{i0} = 1, \text{ for all } i.$$

The relations $(a_1{}^{i}a_2{}^{j})P_{ak}(a_1{}^{i}a_2{}^{j})^{-1}$ merely imply $g_{ij,k} = g_k$ and $\alpha_{ij}g_k = g_k\alpha_{ij}$ for all i, j and for $k \geqq 2$.

Reassuming, we have at present the following generators and generating relations for the group H_n:

Generators of H_n:

$$\alpha_{ij}, g_{ij} \quad (i, j = 0, \pm 1, \pm 2, \cdots); \qquad g_2, \cdots, g_{n-1}.$$

Generating relations of H_n:

---

* The relations $\pi_{ij} = 1$ replace in the present case the relations $T_mF_{kl}T_m{}^{-1}$ given by Reidemeister,[1] p. 13. We use the notations of Reidemeister and we prove that, quite generally, *the ng^2 relations $T_mF_{kl}T_m{}^{-1} = 1$ can be replaced by the g relations $\pi_m(S_{ik}) = 1$*, where $\pi_m(S_{ik})$ is the power product of the S_{ik}'s which we get if we express T_m in the form π_mT_m according to Reidemeister's construction. Let us first consider the case in which $m \neq g$, i. e. T_m is not the element 1. In this case $T_mF_{kl}T_m{}^{-1}$ contains S_{kl}^{-1} and hence (see Reidemeister,[1] p. 11) in the course of the construction S_{kl}^{-1} must be replaced by $T_{g_{ik}}S_k{}^{-1}T_i{}^{-1}$. But then we get an expression which can be changed into $T_mT_m{}^{-1}$ by using the trivial relations $S_iS_i{}^{-1} = 1$. Hence the relation $T_mF_{kl}T_m = 1$, expressed in terms of S_{ik}'s, can be changed into the relation $T_mT_m{}^{-1} = 1$ by using the trivial relations $S_{ik}S_{ik}^{-1} = 1$. Now it is immediately seen that this last relation coincides with the relation $\pi_m{}^{-1}(S_{ik}) = 1$. Let now $m = g$. In this case we have the generating relation $F_{kl} = 1$; but F_{kl} expressed in terms of the S_{ik}'s has the following form:

$$S_{kl}^{-1}\pi_k(S_{ij})S_{kl}\pi_{g_{kl}}^{-1}(S_{ij}),$$

and hence the relations $F_{kl} = 1$ are consequences of the relations $\pi_m = 1$.

† If, for instance, $i \geqq 0$ and $j \geqq 0$, then $\pi_{ij} = \alpha_{00}\alpha_{10} \cdots \alpha_{i-1,0}\alpha'_{i,0}\alpha'_{i1}, \alpha'_{i,j-1}$.

$$\left.\begin{array}{lll}(5) & T^{(ij)}_{kl} = (a_1{}^i a_2{}^j)T_{kl}(a_1{}^i a_2{}^j)^{-1} = 1, & \\ & \qquad (k, l = 1, 2, \cdots, n-1, k \neq l) & \\ (5') & T^{(ij)} = (a_1{}^i a_2{}^j)T \ (a_1{}^i a_j{}^j)^{-1} = 1, & \\ (5'') & S_k{}^{(ij)} = (a_1{}^i a_2{}^j)S_k \ (a_1{}^i a_2{}^j)^{-1} = 1, & \\ & \qquad (k = 1, 2); & \\ (5''') & S^{(ij)}_{12} = (a_1{}^i a_2{}^j)S_{12}(a_1{}^i a_2{}^j)^{-1} = 1 & \end{array}\right\} (i, j = 0, \pm 1, \pm 2, \cdots).$$

$$(5a) \qquad \alpha_{i0} = 1, \qquad (i = 0, \pm 1, \pm 2, \cdots),$$

$$(5b) \qquad \alpha_{ij} g_k = g_k \alpha_{ij}, \quad (k \geq 2).$$

6. The expression of the elements $T^{(ij)}_{kl}$, $T^{(ij)}$, $S_k{}^{(ij)}$, $S^{(ij)}_{12}$ in terms of the generators of H_n leads in a straightforward manner to the following generating relations (where $i, j = 0, \pm 1, \pm 2, \cdots$):

$$\begin{array}{lll}(6) & T^{(ij)}_{1l}: & g_{ij} g_l = g_l g_{ij}, \qquad l > 2; \\ (6') & T^{(ij)}_{kl}: & g_k g_l = g_l g_k; \qquad |l - k| \neq 1;\ k, l > 1; \\ (6'') & T^{(ij)}_{12}: & g_{ij} g_2 g_{ij} = g_2 g_{ij} g_2; \\ (6''') & T^{(ij)}_{k,k+1}: & g_k g_{k+1} g_k = g_{k+1} g_k g_{k+1}, \qquad k > 1; \\ (7) & T^{(ij)}: & g_{ij} g_2 \cdots g_{n-2} g^2_{n-1} g_{n-2} \cdots g_2 g_{ij} = \alpha_{ij} \alpha^{-1}_{i,j-1}; \\ (8_1) & S_1{}^{(ij)}: & g_{i+2,j} = (\alpha_{ij} g^{-1}_{i+1,j} \alpha_{i+1,j})^{-1} g_{i,j} (\alpha_{ij} g^{-1}_{i+1,j} \alpha_{i+1,j}); \\ (8_2) & S_2{}^{(ij)}: & g_{i,j+2} = g_{i,j+1} g_{i,j} g^{-1}_{i,j+1}; \\ (9) & S^{(ij)}_{12}: & \alpha_{ij} g_{i+1,j} g_{i+1,j-1} \alpha^{-1}_{i,j-1} g_{i,j-1} g^{-1}_{ij} = 1.\end{array}$$

The relations (8_2) imply that $g_{i,j+1} g_{i,j}$ is independent of j. Let for brevity,

$$(10) \qquad g_{i,j+1} g_{i,j} = s_i.$$

The recurrence relations (9) allow us to express all α_{ij}'s in terms of the $g_{\alpha\beta}$'s, $\alpha_{i,0}$. Taking into account (5a) and (10) we find

$$(9') \qquad \alpha_{ij} = g_{ij} g^{-1}_{i0} \, s^{-j}_{i+1}, \qquad (i, j = 0, \pm 1, \pm 2, \cdots).$$

Substituting these expressions of the α_{ij}'s into the relations (7) and taking into account (10) we find in a straightforward manner that the relations (7) can be replaced by the following relations:

$$(7') \quad g_{i1} g_{i+1,1} g_{i+1,0} g_{i,0} g_2 \cdots g_{n-2} g^2_{n-1} g_{n-2} \cdots g_2 = 1, \quad (i = 0, \pm 1, \pm 2, \cdots).$$

We have thus obtained a first reduction of the algebraic expression of the group H_n: *as generators of* H_n *we have the elements* g_{ij}, $(i, j = 0, \pm 1, \pm 2, \cdots)$, $g_2, \cdots, g_{n-1}$; *the generating relations are* (5b), (6)-(6'''), (7'), (8_1), (8_2), *where the elements* α_{ij} *in* (8_1) *are defined by* (9'), *and* (10).

Since $\alpha_{i0} = 1$, the relation (8_1) for $j = 0$ yields the following relation:

$$(11) \qquad g_{i+1,0}g_{i,0} = g_{i,0}g_{i-1,0}, \qquad (i = 0, \pm 1, \pm 2, \cdots).$$

Since, by (7′), the product $g_{i,1}g_{i+1,1}g_{i+1,0}g_{i,0}$ is independent of i, we deduce, as a consequence of (11), the following relation:

$$(11') \qquad g_{i,1}g_{i+1,1} = g_{i-1,1}g_{i,1}, \qquad (i = 0, \pm 1, \pm 2, \cdots).$$

We proceed to prove that *in the reduced complete set of generating relations of H_n given above, the infinite set of relations* (8_1) *can be replaced by the relations* (11) *and* (11′), *i. e. the relations* (8_1) *are group-theoretic consequences of the remaining generating relations and of* (11) *and* (11′).

Proof. We denote by τ the product $g_2 \cdots g_{n-2}g^2_{n-1}g_{n-2} \cdots g_2$, or, in view of (7′),

$$(12) \qquad \tau = (g_{i1}g_{i+1,1}g_{i+1,0}g_{i,0})^{-1}.$$

We have

$$\tau^{-1} = g_{i1}s_{i+1}g_{i0} = s_i g^{-1}_{i0} s_{i+1} g_{i0},$$

hence

$$s^{-1}_{i+1} = g_{i,0}\tau s_i g^{-1}_{i,0}.$$

Substituting into (9′) we get

$$\alpha_{ij} = g_{ij}(\tau s_i)^j g^{-1}_{i,0},$$

and hence, using a new symbol β_{ij} for the transforming elements in (8_1), we have

$$(13) \qquad \beta_{ij} = \alpha_{ij}g^{-1}_{i+1,j}\alpha_{i+1,j} = g_{ij}(\tau s_i)^j g^{-1}_{i0}(\tau s_{i+1})^j g^{-1}_{i+1,0}.$$

By (5b) each element α_{ij} is commutative with τ, since $\tau = g_2 \cdots g^2_{n-1} \cdots g_2$, hence τ is also commutative with $\alpha_{ij}\alpha^{-1}_{i,j-1}$, i. e. in view of (7) (which is a consequence of (7′) and (9′)), τ is commutative with $g_{ij}\tau g_{ij}$:

$$(14) \qquad (\tau g_{ij})^2 = (g_{ij}\tau)^2.$$

We use the following relations:

$$(15) \qquad (\tau s_i)g^{-1}_{i0} = \tau g_{i1} \quad [\text{by } (10)]$$

$$(15') \qquad \tau s_i(\tau g_{i1})(\tau s_{i+1}) = \tau g_{12}\tau g_{i1}g^{-1}_{i0} \quad [\text{by } (10), (12) \text{ and } (14)]$$

$$(15'') \qquad \tau s_i(\tau g_{ij}\tau g_{i,j-1}) = (\tau g_{i,j+1}\tau g_{ij})\tau s_i \quad [\text{by } (10) \text{ and } (14)].$$

To prove (15′) and (15″) we proceed as follows:

$$s_i\tau g_{ij}\tau = g_{i,j+1}g_{ij}\tau g_{ij}\tau \quad [\text{by } (10)] = g_{i,j+1}\tau g_{ij}\tau g_{ij};$$

hence

$$\tau s_i(\tau g_{i1})(\tau s_{i+1}) = \tau g_{i2}\tau g_{i1}\tau g_{i1}s_{i+1} = \tau g_{i2}\tau g_{i1}g^{-1}_{i0} \quad [\text{by } (12)],$$

and

$$\tau s_i(\tau g_{ij}\tau g_{i,j-1}) = \tau g_{i,j+1}\tau g_{ij}\tau g_{ij} g_{i,j-1} = \tau g_i$$

From (15), (15′), and (15″) we deduce for any integ
relations:

$$(16) \qquad \gamma_{i,2k} = (\tau s_i)^{2k}\, g_{i,0}^{-1}(\tau s_{i+1})^k = \tau g_{i,2k} \cdots \tau g_i$$

$$(16') \qquad \gamma_{i,2k+1} = (\tau s_i)^{2k+1} g_{i,0}^{-1}(\tau s_{i+1})^k = \tau g_{i,2k+1} \cdots \tau g_i$$

In an exactly similar manner the following relations ca

$$(\tau s_{i+1}) g_{i0}(\tau s_i) = \tau g_{i0};$$
$$\tau g_{i0} \cdot \tau s_i = g_{i0} \cdot \tau g_{i0}\tau g_{i,-1}.$$

From these relations and from (15″) we deduce for
following relations:

$$(17) \qquad \gamma_{i,2k} = (\tau s_i)^{2k}\, g_{i0}^{-1}(\tau s_{i+1})^k = g_{i,2k+1}^{-1}\tau^{-1}g_{i,2k+2}^{-1}\tau^{-}$$

$$(17') \qquad \gamma_{i,2k+1} = (\tau s_i)^{2k+1} g_{i0}^{-1}(\tau s_{i+1})^k = g_{i,2k+2}^{-1}\tau^{-1} \cdots g_{i0}^{-1}$$

Using the relations (16), (16′), (17) and (17′) and t
relations ($\mathcal{E}_2$) and (14), we obtain easily the following

$$\left.\begin{array}{lr} (18) & \gamma_{i,2k}^{-1}\, g_{i,2k}\, \gamma_{i,2k} = g_{i0} \\ (18') & \gamma_{i,2k+1}^{-1} g_{i,2k+1}\gamma_{i,2k+1} = \tau g_{i1}\tau^{-1} \end{array}\right\} k \gtreqless$$

We put

$$(\tau s_{i+1})^k g_{i+1,0}^{-1} = \delta_{i,k}$$

so that, by (13),

$$\beta_{i,2k} = g_{i,2k}\, \gamma_{i,2k}\, \delta_{i,k},$$
$$\beta_{i,2k+1} = g_{i,2k+1}\gamma_{i,2k+1}\delta_{i,k+1}.$$

In view of (18) and (18′), the relations ($\mathcal{E}_1$) will follo
lish the following relations:

$$(19) \qquad \delta_{ik}^{-1} g_{i0}\delta_{ik} = g_{i+2,2k},$$

$$(19') \qquad \delta_{ik}^{-1}\tau g_{i1}\tau^{-1}\delta_{ik} = g_{i+2,2k-1}.$$

Since $\delta_{i0} = g_{i+1,0}^{-1}$ and $\tau^{-1}\delta_{i1} = s_{i+1}g_{i+0.0}^{-1} = g_{i+1,1}$, the re
coincides with (11) and the relations (19′) for $k = 1$
Hence in order to establish the relations (19) and (19′)
to show that if they hold true for a given k, they also ho
for $k - 1$. Now

$$\delta_{i,k+1} = \delta_{ik} g_{i+1,0}\tau s_{i+1} g_{i+1,0}^{-1} = \delta_{ik} g_{i+1,0}(g_{i+1,1}g_{i+2,1}g_{i+}$$
$$= \delta_{ik}(g_{i+2,1}g_{i+2,0})^{-1} = \delta_{ik}s_{i+2}^{-1}.$$

Hence, assuming (19) and (19′) for a given value of k, the same relations follow also for $k+1$ and $k-1$ in view of the relations

$$s_{i+2}^{-\epsilon} g_{i+2,j} s_{i+2}^{\epsilon} = g_{i+2,j-2\epsilon}, \qquad \epsilon = \pm 1$$

which are direct consequences of the relations (8_2), q. e. d.

7. We now complete the elimination of the elements α_{ij} from the generating relations of H_n by proving that also the commutativity relations $\alpha_{ij} g_k = g_k \alpha_{ij}$, $k > 2$, (5b), are consequences of the remaining relations, to wit, of the relations (6), (6″), (8_2), and (7′). This is obvious if $k > 3$, since the α_{ij} depend only on the g_{ij}'s [see (9′)] and since, by (6), the g_{ij}'s are commutative with $g_3, \cdots, g_{n-1}$. By (7), which is a consequence of (10) and of the relations (9′) which define the elements α_{ij}, we have

$$\alpha_{ij} \alpha_{i,j-1}^{-1} = g_{ij} g_2 \cdots g^2_{n-1} \cdots g_2 g_{ij}.$$

Since $\alpha_{i0} = 1$, it is sufficient to establish the commutativity of $\alpha_{ij}\alpha_{i,j-1}^{-1}$ and g_2. Now, using the relations (6) and (6‴) we find:

$$\begin{aligned} g_2 \cdot g_{ij} g_2 g_3 \cdots g^2_{n-1} \cdots g_3 g_2 g_{ij} &= g_{ij} g_2 g_{ij} g_3 \cdots g^2_{n-1} \cdots g_3 g_2 g_{ij} \\ &= g_{ij} g_2 \cdots g^2_{n-1} \cdots g_3 g_{ij} g_2 g_{ij} = g_{ij} g_2 \cdots g^2_{n-1} \cdots g_2 g_{ij} g_2, \end{aligned}$$

and this proves our assertion.

Reassuming the reduction carried out so far, we have that our group H_n is defined by the following set of generators and generating relations:

Generators:

(20) $\qquad g_{ij} \ (i, j = 0, \pm 1, \pm 2, \cdots), \qquad g_2, \cdots, g_{n-1}.$

Generating relations:

(21) $\qquad g_{i,j+2} = g_{i,j+1} g_{ij} g_{i,j+1}^{-1}$

(22) $\qquad g_{i+2,0} = g_{i+1,0} g_{i,0} g_{i+1,0}^{-1}$

(22_1) $\qquad g_{i+2,1} = g_{i+1,1}^{-1} g_{i,1} g_{i+1,1}$

(23) $\qquad g_{ij} g_2 g_{ij} = g_2 g_{ij} g_2$

(23_1) $\qquad g_{ij} g_k = g_k g_{ij} \qquad (k = 3, \cdots, n-1)$

(24) $\qquad g_k g_{k+1} g_k = g_{k+1} g_k g_{k+1} \qquad (k = 2, 3, \cdots, n-2)$

(24_1) $\qquad g_k g_l = g_l g_k, \qquad |k - l| \neq 1$

(25) $\qquad g_{i1} g_{i+1,1} g_{i+1,0} g_{i,0} g_2 \cdots g_{n-2} g^2_{n-1} g_{n-2} \cdots g_2 = 1.$

The existence of a finite set of generators follows now readily. In fact, the relations (21), for a fixed value of i, can be considered as recurrence relations which define the elements g_{ij} in terms of the two free elements g_{i0} and g_{i1}. Then the relations (22) and (22_1) can be used in order to express all the elements g_{i0} and g_{i1} in terms of g_{00}, g_{10} and g_{01}, g_{11}, respectively. *Consequently our group H_n is generated by the $n+2$ elements:*

$$g_{00}, g_{10}, g_{01}, g_{11}, g_2, \cdots, g_{n-1}. \tag{26}$$

The reduction of the infinite set of relations (23_1) is trivial: since all the g_{ij}'s are expressible in terms of g_{00}, g_{10}, g_{01}, g_{11}, all the relation (23_1) are consequences of the four relations

$$g_{ij}g_k = g_k g_{ij}, \qquad (i, j = 0, 1). \tag{23_2}$$

The reduction of the relations (23) is based on the following

LEMMA. *If four elements a, b, c, x satisfy the relations*

$$axa = xax, \quad bxb = xbx, \quad cxc = xcx, \quad c = b^{-\epsilon}ab^{\epsilon}, \quad \epsilon = 1 \text{ or } -1,$$

and if $d = c^{-\epsilon}bc^{\epsilon}$, then the above relations have as a group-theoretic consequence the relation $xdx = dxd$.

Proof. Let $\epsilon = +1$. Then

$$\begin{aligned} bcxc^{-1}b^{-1} &= bx^{-1}cxb^{-1} = bx^{-1}b^{-1}abxb^{-1} = x^{-1}b^{-1}xax^{-1}bx \\ &= x^{-1}b^{-1}a^{-1}xabx = x^{-1}c^{-1}b^{-1}xbcx. \end{aligned}$$

Hence

$$\begin{aligned} dxd^{-1} &= c^{-1}bcxc^{-1}b^{-1}c = x^{-1}c^{-1}x^{-1}b^{-1}xbxcx = x^{-1}c^{-1}x^{-1}xbx^{-1}xcx \\ &= x^{-1}c^{-1}bcx = x^{-1}dx. \end{aligned}$$

The case $\epsilon = -1$ is reducible to the case $\epsilon = +1$, by the substitution

$$x_1 = x^{-1}, \quad a_1 = a^{-1}, \quad b_1 = b^{-1}; \quad c_1 = c^{-1}.$$

Putting $a = g_{ij}$, $b = g_{i,j+1}$, $c = g_{i,j+2}$, $\epsilon = -1$, we deduce from the above lemma, in view of (21), that for a fixed i, the relations (23) relative to the indices j, $j+1$, $j+2$ imply as a group-theoretic consequence the relation (23) for the index $j+3$. Similarly, if we put $a = g_{i,j+2}$, $b = g_{i,j+1}$, $c = g_{ij}$, $\epsilon = 1$, we find that the just mentioned three consecutive relations (23) also imply the relation (23) for the index $j-1$. It follows, that for a fixed i, all the relations (23) are consequences of any three of them relative to three consecutive indices, say $j = 0, 1, 2$:

$$g_{ij}g_2g_{ij} = g_2g_{ij}g_2, \qquad (j = 0, 1, 2). \tag{27}$$

Now, in view of (22) and (22_1), we conclude in a similar manner, on the basis of the preceding lemma, that for $j = 0, 1$ the relations $g_{ij}g_2g_{ij} = g_2g_{ij}g_2$ are consequences of three of these relations relative to three consecutive valus of i, say $i = 0, 1, 2$. It remains to consider the set of relations $g_{i2}g_2g_{i2} = g_2g_{i2}g_2$. Using the relations (27) for $j = 0, 1$ and the expression of g_{i2} derived from (21), for $j = 0$, we change the above relation into an equivalent relation as follows:

$$\begin{aligned} g_{i2}g_2g_{i2}g_2^{-1}g_{i2}^{-1}g_2^{-1} &= g_{i1}g_{i0}\underline{g_{i1}^{-1}g_2g_{i1}}g_{i0}\underline{g_{i1}^{-1}g_2^{-1}g_{i1}}g_{i0}^{-1}g_{i1}^{-1}g_2^{-1} \\ &= g_{i1}g_{i0}g_2g_{i1}\underline{g_2^{-1}g_{i0}g_2}g_{i1}^{-1}g_2^{-1}g_{i0}^{-1}g_{i1}^{-1}g_2^{-1} = (g_{i1}g_{i0}g_2)^2(g_2g_{i1}g_{i0})^{-2}. \end{aligned}$$

Hence the relation $g_{i2}g_2g_{i2} = g_2g_{i2}g_2$ can be replaced by the relation

$$(g_{i1}g_{i0}g_2)^2 = (g_2g_{i1}g_{i0})^2. \tag{28}$$

Now, it is not difficult to see that the expressions $\sigma_i = (g_{i1}g_{i0}g_2)^2(g_2g_{i1}g_{i0})^{-2}$ are all transforms of each other, for $i = 0, \pm 1, \pm 2, \cdots$, as a consequence of the relations (27) ($j = 0, 1$), (23_1) and (25). In fact, let

$$g_3 \cdots g_{n-2}g^2_{n-1}g_{n-2} \cdots g_3 = \delta,$$

so that, by (23_1), we have $g_{ij}\delta = \delta g_{ij}$. By (25), we have

$$g_{i+1,1}g_{i+1,0} = (g_{i0}g_2\delta g_2g_{i1})^{-1}.$$

Hence, substituting into σ_{i+1} we find

$$\begin{aligned} \sigma_{i+1} &= (g_{i1}^{-1}g_2^{-1}\delta^{-1}g_2^{-1}g_{i0}^{-1}g_2)^2(g_{i0}g_2\delta g_2g_{i1}g_2^{-1})^2 \\ &= g_{i1}^{-1}g_2^{-1}\delta^{-1}g_2^{-1}g_{i0}^{-1}\underline{g_2g_{i1}^{-1}g_2^{-1}\delta^{-1}g_{i0}\delta}g_2g_{i1}\underline{g_2^{-1}g_{i0}g_2\delta g_2g_{i1}g_2^{-1}} \\ &= g_{i1}^{-1}g_2^{-1}\delta^{-1}g_2^{-1}(g_2g_{i1}g_{i0})^{-1}(g_{i1}g_{i0}g_2)^2g_{i0}^{-1}g_{i1}^{-1}\delta g_2g_{i1} \\ &= (g_{i0}^{-1}g_{i1}^{-1}\delta g_2g_{i1})^{-1}\bar{\sigma}_i(g_{i0}^{-1}g_{i1}^{-1}\delta g_2g_{i1}), \end{aligned}$$

where $\bar{\sigma}_i = (g_2g_{i1}g_{i0})^{-2}(g_{i1}g_{i0}g_2)^2$, obviously a transform of σ_i.

Hence, any one of the relations (28) implies as a consequence the entire set of these relations. We shall take the relations relative to $i = 0$.

We finally observe that in view of the relations $g_{i1}g_{i+1,1} = g_{i+1,1}g_{i+2,1}$ and $g_{i+1,0}g_{i,0} = g_{i+2,0}g_{i+1,0}$ [(22_1) and (22) respectively], the infinite set of relations (25) reduces to one relation, say relative to $i = 0$.

Reassuming, we have the following result:

The group H_n is defined by the following set of generators and generating relations:

1. *Generators:*

(29) $$g_{00}, g_{10}, g_{01}, g_{11}, g_2, \cdots, g_{n-1}$$

2. *Generating relations:*

(30) $$g_{ij}g_2g_{ij} = g_2g_{ij}g_2, \qquad (i, j = 0, 1).$$

(30′) $$(g_{10}g_{00}g_2)^2 = (g_2g_{10}g_{00})^2$$

(30″) $$(g_{01}g_{11}g_2)^2 = (g_2g_{01}g_{11})^2$$

(30‴) $$(g_{01}g_{00}g_2)^2 = (g_2g_{01}g_{00})^2$$

(31) $$g_{ij}g_k = g_kg_{ij}, \qquad k > 2$$

(32) $$g_kg_{k+1}g_k = g_{k+1}g_kg_{k+1}, \qquad (k = 2, 3, \cdots, n-2)$$

(32′) $$g_kg_l = g_lg_k, \qquad |k - l| \neq 1$$

(33) $$g_{01}g_{11}g_{10}g_{00}g_2 \cdots g_{n-2}g^2_{n-1}g_{n-2} \cdots g_2 = 1.$$

The relations (30′), (30″) and (30‴) are the relations (27) relative to the following values of the indices i, j: $i = 2$, $j = 0$; $i = 2$, $j = 1$; $i = 0$, $j = 2$, after the expressions of g_{20}, g_{21}, g_{02}, given by (22), (22_1) and (21) respectively, are substituted.

Remark. If we change our notation as follows:

(34) $$g_{01} = \lambda_1, \quad g_{11} = \lambda_2, \quad g_{10} = \lambda_3, \quad g_{00} = \lambda_4,$$

then we see that the relations (30′), (30″), (30‴) are of the form

(35) $$(g_2\lambda_i\lambda_j)^2 = (\lambda_i\lambda_jg_2)^2, \quad i < j, \quad i, j = 1, 2, 3, 4.$$

An easy verification shows that *the six relations* (35) *all hold true.* For $i = 3$, $j = 4$; $i = 1$, $j = 2$; $i = 1$, $j = 4$, they coincide with the relations (30′), (30″), (30‴) respectively. For $i = 2$, $j = 3$ the relation (35) coincides with (28) for $i = 1$. The reader will easily verify the truth of the relations (35) in the remaining two cases.

III. The fundamental group of plane elliptic curves.

8. What is the geometric significance of the invariant subgroup H_n of G_n? We proceed to show that H_n is the fundamental group of a certain subspace of $R^n - D$ (see section 1).

Let us consider R as the Riemann surface of some algebraic elliptic curve f. This curve carries, for every n, a simple infinity of complete linear series g_n^{n-1} of dimension $n-1$, and each set of n points of f belongs to one and only

one series g_n^{n-1}. The simple infinity of these series is an elliptic one-dimensional variety, birationally equivalent to f: in fact, each series contains a unique set of n points of which $n-1$ are preassigned fixed points $P_1^0, \cdots, P^0_{n-1}$, and then the n-th point P of the set determines the series uniquely.

Since a g_n^{n-1} is represented on R^n by a space homeomorphic to a linear space of $n-1$ dimensions, we conclude that *the algebraic variety R^n contains an elliptic pencil $\{S_{n-1}\}$ of linear $(n-1)$-spaces S_{n-1} free from base points.*

Let V_{n-2} be the intersection of an S_{n-1} with the discriminant variety D of R^n. We assert that *H_n is the fundamental group of the residual space $S_{n-1} - V_{n-2}$.*

Proof. We take the origin O of the fundamental group to be a point of $S_{n-1} - V_{n-2}$. We have to prove that: (1) a singular 1-sphere γ on O and in $R^n - D$ represents an element of H_n, if and only if it can be deformed over $R^n - D$ into a 1-sphere γ' contained in $S_{n-1} - V_{n-2}$, the point O being fixed; (2) if γ is already in $S_{n-1} - V_{n-2}$ and if it bounds a singular 2-cell on $R^n - D$, then it also bounds a singular 2-cell on $S_{n-1} - V_{n-2}$.

Let u be an elliptic integral of the first kind attached to the curve f. It is well known that f admits a continuous one-parameter group of birational transformations π_t into itself, represented analytically by the equation $u' \equiv u + t$ (mod. periods). Each transformation π_t of the group is an automorphism of R. There is an induced automorphism of R^n, which we shall also denote by π_t and which is at the same time an automorphism of the residual space $R^n - D$, since π_t transforms sets of n distinct points of R into sets of n distinct points. Each π_t permutes the linear spaces S_{n-1} (images of linear series on f) and induces a birational transformation σ_τ into itself (an automorphism) of the elliptic pencil $\{S_{n-1}\}$. If we put

$$v = u(x_1) + u(x_2) + \cdots + u(x_n),$$

where $x_1, \cdots, x_n$ is an n-tuple of points of R, then v is a simple integral attached to R^n and v reduces to a constant on each member of the pencil S_{n-1} (theorem of Abel). Hence v is also an elliptic integral of the first kind attached to the pencil $\{S_{n-1}\}$, and the transformation σ_τ is given by the equation

$$v' \equiv v + \tau, \qquad \tau \equiv nt.$$

The group of the transformations π_t covers n^2 times the group of transformations σ_τ, since to the identity σ_0 correspond the n^2 transformations $\pi_{\omega/n}$, where ω/n is the n-th of a period of u. Since the covering is free from branch points, it follows immediately that *any variation of S_{n-1} in the pencil $\{S_{n-1}\}$ can be*

accompanied by an isotopic deformation of the variable S_{n-1}, *and in such a manner that also the residual space is deformed isotopically.* This isotopic deformation of $S_{n-1} - V_{n-2}$ is simply effected by a convenient chain of transformations π_t, $0 \leq t \leq 1$, applied to the initial position of the S_{n-1}. From this last statement we derive immediately the following conclusion. Let R' be the Riemann surface of the elliptic pencil $\{S_{n-1}\}$. Every point P of R^n lies on a definite S_{n-1} of the pencil and is thus mapped upon a definite point P' of R'. Similarly any point set A on R^n is mapped upon a point set A' of R'. A deformation of A on R^n or on $R^n - D$ induces a deformation of A' on R'. From the preceding statement we may conclude that, conversely, *if A is on $R^n - D$, then any deformation of A' on R' is induced by a deformation of A on $R^n - D$, and that if a point P of A' is fixed throughout the deformation of A', then the points of A which are mapped upon P may also be assumed to be fixed during the deformation of A.*

Let γ be a singular 1-sphere on $R^n - D$ issued for the origin O of the group G_n. From the definition of the group H_n follows that γ represents an element of H_n if and only if the map γ' of γ on R' is a (singular) 1-cycle ~ 0. Assume that $\gamma' \sim 0$. Then γ' can be contracted on R' to the point O', image of O, and hence, by our preceding result, γ can be deformed into a 1-sphere γ_1 contained in $S_{n-1} - V_{n-2}$, the point O being fixed. Conversely, if γ can be deformed into such a 1-sphere γ_1, then γ' can be contracted to the point O', $\gamma' \sim 0$, and hence γ represents an element of H_n.

Assume that γ is in $S_{n-1} - V_{n-2}$ and that it bounds a (singular) 2-cell E_2 on $R^n - D$. The map of E_2 on R' is a 2-sphere M' containing O', since the boundary of E_2 is mapped on the point O'. On R' any 2-sphere on O' can be contracted to the point O', this last point being fixed. Hence E_2 can be deformed on $R^n - D$ into a 2-cell contained in $S_{n-1} - V_{n-2}$, the boundary γ being fixed. This completes the proof of our theorem.

9. The variety V_{n-2} is an hypersurface immersed in the $(n-1)$-space S_{n-1}. Let S_2 be a general plane of S_{n-1} and let C be the plane algebraic curve along which S_2 cuts V_{n-2}. By a theorem proved in Zariski,[5] the fundamental group H_n of the residual space $S_{n-1} - V_{n-2}$ coincides with the fundamental group of $S_2 - C$. Now, a general plane S_2 in our S_{n-1} is the image of a general series g_n^2 immersed in the corresponding g_n^{n-1} of the elliptic curve. A point of S_2 represents a set of the series g_n^2. If we refer the sets of the g_n^2 to the lines of a plane, we obtain the general plane elliptic curve Γ, of order n, a birational transform of f, on which the sets of the g_n^2 are cut out by the lines of the plane. The points of S_2 which are on C represent n-tuples of the g_n^2

with coincident points, hence correspond to the tangent lines of Γ. We conclude that *the plane curve C is the dual of a general elliptic plane curve Γ of order n, and that our group H_n is the fundamental group of the residual space of C.*

The curve C is of order $2n$, possesses $k = 3n$ cusps and $d = 2n(n-3)$ nodes, and is the maximal cuspidal elliptic curve of its order. Of the generating relations of H_n, the relations (30), (30′), (30″), (30‴), and (32) are all of the form $aba = bab$ and arise from the cusps of C. The commutativity relations (31) and (32′) are due to the nodes of C. Finally, the relations (33) corresponds to the relation $\lambda_1\lambda_2 \cdots \lambda_{2n} = 1$, where the λ_i's are loops contained in a general line of the plane of C and surrounding the $2n$ intersections of this line with C (see Zariski [2]). The relations $\lambda_5 = \lambda_{2n}$, $\lambda_6 = \lambda_{2n-1}$ etc., arise from the n tangent lines of C belonging to a pencil of lines (compare the analogous discussion of the singularities and of the corresponding generating relations in Zariski,[4] p. 615). By watching the effect which the removal of a cusp or of a node has upon the fundamental group, we arrive at conclusions relative to the fundamental group of any plane curve C' which admits C as a limiting case, in particular of any plane elliptic curve of even order $2n$, possessing only nodes and cusps (compare Zariski,[4] p. 616). Let us first remove a node, i. e. let us consider a node of C as virtually nonexistent. For the fundamental group this amounts to replacing a commutativity relation $ab = ba$ by the relation $a = b$. We may assume that the relation thus affected is the relation $g_2g_4 = g_4g_2$. We have then $g_4 = g_2$. Since $g_2g_5 = g_5g_2$ and $g_4g_5g_4 = g_5g_4g_5$, it follows $g_4 = g_5$. In a similar manner we find $g_2 = g_3 = g_4 = \cdots = g_{n-1}$. Since $g_{ij}g_2g_{ij} = g_2g_{ij}g_2$ and $g_{ij}g_3 = g_3g_{ij}$, the relation $g_2 = g_3$ implies the relation $g_{ij} = g_2$. Hence the fundamental group becomes a cyclic group of order $2n$. Let us now remove a cusp, by converting the cusp into a node. A relation $aba = bab$ will be affected and will have to be replaced by $a = b$. We may assume that the affected relation is the relation $g_{00}g_2g_{00} = g_2g_{00}g_2$. We have then $g_2 = g_{00}$, after the cusp has been removed. If $n \geq 4$, we may use the relations $g_{00}g_3 = g_3g_{00}$, $g_2g_3g_2 = g_3g_2g_3$ and we then find that $g_2 = g_3$. We conclude as before that the group becomes cyclic. Hence, we have the following result: *If C' is a plane curve of order $2n > 6$ with nodes and cusps, and if C' admits the maximal cuspidal elliptic curve C, of the same order, as a limiting case, without being a curve C itself, then the fundamental group of C' is cyclic (of order $2n$). In particular, every plane elliptic curve of even order $2n$ possessing less than $3n$ cusps has a cyclic fundamental group.*

In the exceptional case $n = 3$ we are dealing with the dual of a general

plane cubic, i. e. with an elliptic sextic having 9 cusps. We write the generating relations of H_3, using (27) instead of the equivalent set of relations (30)-(30‴):

$$g_{ij}g_2g_{ij} = g_2g_{ij}g_2, \qquad (i, j = 0, 1, 2). \tag{34}$$

$$g_{01}g_{11}g_{10}g_{00}g_2{}^2 = 1, \tag{35}$$

where

$$\begin{cases} g_{i2} = g_{i1}g_{i0}g_{i1}^{-1} \\ g_{20} = g_{10}g_{00}g_{10}^{-1} \\ g_{21} = g_{11}^{-1}g_{01}g_{11}. \end{cases} \tag{36}$$

The 9 relations (34) are typical cuspidal relations, and one may conjecture that they correspond to the 9 cusps of the curve. However, since only 7 of these relations are group-theoretically independent, this conjecture requires proof. The 7 independent relations are given by (30), (30′), (30″), (30‴) and correspond to the following values of i, j:

$$i, j = 0, 1; \quad i = 2, j = 0; \quad i = 2, j = 1; \quad i = 0, j = 2.$$

At present we can only assert that the seven independent relations are relations at 7 of the cusps. We recall that the well known group of 9 flexes of a cubic curve is doubly transitive. Hence if we remove a certain number of cusps of C, it is immaterial which cusps are removed, as long as the number of removed cusps does not exceed 2.

We remove successively the two cusps which give rise to the relations $g_{00}g_2g_{00} = g_2g_{00}g_2$, $g_{10}g_2g_{10} = g_2g_{10}g_2$. After the removal of the first cusp we have $g_{00} = g_2$. The relations (30′) and (30‴) become then consequences of the relations (30) for $i = 1$, $j = 0$ and $i = 0$, $j = 1$ respectively, while the relation (30″) becomes a consequence of (33). Hence the fundamental group of a sextic with 8 cusps (and with one or no double points) is generated by 4 elements

$$g_{10}, g_{01}, g_{11}, g_2$$

satisfying the relations:

$$g_{ij}g_2g_{ij} = g_2g_{ij}g_2$$
$$g_{01}g_{11}g_{10}g_2{}^3 = 1.$$

If we now remove the second cusp, we get $g_{10} = g_2$, and thus the fundamental group of a sextic with 7 cusps and of genus ≥ 1 is generated by 3 elements

$$g_{01}, g_{11}, g_2$$

satisfying the 3 relations:

$$g_{01}g_2g_{01} = g_2g_{01}g_2$$
$$g_{11}g_2g_{11} = g_2g_{11}g_2$$
$$g_{01}g_{11}g_2^4 = 1.$$

This is also a group generated by the two elements

$$u = g_{11}g_2g_{11}, \quad v = g_{11}g_2$$

satisfying the relations $u^2 = v^3 = 1$.

It is known that this group is also the fundamental group of a sextic with six cusps on a conic (Zariski [2]). Hence there must be among the seven cusps left, a third cusp whose removal has no effect on the fundamental group. It is easily seen, by using the relations (36), that if the removal of an additional cusp yield the relation $g_{ij} = g_2$, $i \neq 2$, $j \neq 0$, the group becomes cyclic. On the contrary, if the removed cusp gave rise originally to the relation $g_{20}g_2g_{20} = g_2g_{20}g_2$, the group is unaltered, since the relation $g_{20} = g_2$ is already implied, in view of (36), by the removal of the first two cusps ($g_{00} = g_{10} = g_2$).

It is known that the sextics with six cusps distribute themselves into two distinct continuous systems, according as the six cusps lie or do not lie on a conic. The preceding considerations lead therefore to the conclusion that *the fundamental group of a sextic with six cusps not on a conic is cyclic* (of period 6).

One can verify the following: if $g_{i_1j_1}, g_{i_2j_2}, g_{i_3j_3}$ are any 3 of our nine elements g_{ij} such that $i_1 + i_2 + i_3 \equiv j_1 + j_2 + j_3 \equiv 0(3)$, then the removal of the three corresponding cusps (whence the addition of the relations $g_{i_1j_1} = g_{i_2j_2} = g_{i_3j_3} = g_2$) leads to a curve whose fundamental group is the above mentioned group of a sextic with six cusps on a conic. If, however, the above congruences do not hold true simultaneously, then the removal of the corresponding cusps leads to a curve with a cyclic fundamental group. What we have here is obviously something which adds topological significance to the configuration of the 12 MacLaurin lines determined by the nine flexes of a cubic curve. It is known that if the nine flexes are distributed into three triples lying on three MacLaurin lines, then the six flex tangents of any two of the triples lie on a line conic. Dually, any two of the corresponding triples of cusps lie on a conic. If then the three cusps of one triple are considered as virtual non-existent, the resulting sextic must have six cusps on a conic.

This proves incidentally, that the nine relations (34) reproduce exactly the relations at the nine cusps.

10. We conclude by pointing out that the reasoning employed in our paper,[4] section 7, can be applied also in the present case to elliptic curves of odd order and leads to the conclusion that the fundamental group of such a curve (with nodes and cusps) is always cyclic. For the proof it is sufficient to consider the maximal cuspidal elliptic curve C_{2n+1}, of order $2n+1$, and to observe that C_{2n+1} can be degenerated into the maximal cuspidal elliptic curve C_{2n} and into a line p tangent to C_{2n}. The fundamental group of $C_{2n}+p$ can be obtained from the fundamental group H_n of C_{2n} by adding an extra generator γ and the relations $(\gamma g_2)^2=(g_2\gamma)^2$, $\gamma g_{ij}=g_{ij}\gamma$, $\gamma g_k=g_k\gamma$, $k>2$. We obtain the curve C_{2n+1} by considering the tacnode of $C_{2n}+p$ at the point of tangency of p, as a virtual cusp. As a consequence, we replace the relation $(\gamma g_2)^2=(g_2\gamma)^2$ by the relation $g_2=\gamma$, and from this follows immediately that the group of C_{2n+1} is cyclic.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] K. Reidemeister, "Knoten und Gruppen," *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität*, vol. 5 (1927).

[2] O. Zariski, "On the problem of existence of algebraic functions of two variables possessing a given branch curve," *American Journal of Mathematics*, vol. 51 (1929).

[3] O. Zariski, "A topological proof of the Riemann-Roch theorem on an algebraic curve," *American Journal of Mathematics*, vol. 58 (1936).

[4] O. Zariski, "On the Poincaré group of rational plane curves," *American Journal of Mathematics*, vol. 58 (1936).

[5] O. Zariski, "A theorem on the Poincaré group of an algebraic hypersurface," *Annals of Mathematics*, vol. 38 (1937).

ON THOSE POINTS OF AN ALGEBRAIC MANIFOLD NOT REACHABLE BY A GIVEN PARAMETRIC REPRESENTATION.[1]

By J. F. DALY.

Let K denote the complex field, and let $x_1, \cdots, x_n$ be elements of an algebraic extension $K(t_1, \cdots, t_m, x_1, \cdots, x_n)$ of the pure transcendental field $K(t_1, \cdots, t_m)$. A set of complex numbers $t'_1, \cdots, t'_m, x'_1, \cdots, x'_n$ will be called allowable if every polynomial $F(t, x)$ which vanishes as an element of $K(t_1, \cdots, t_m, x_1, \cdots, x_n)$ vanishes also when t'_i, x'_i are substituted for t_i, x_i respectively. The totality of points whose coördinates $\{x'_1, \cdots, x'_n\}$ belong to allowable sets will be contained in some smallest algebraic manifold $\mathfrak{M}$. This manifold is said to be represented parametrically in terms of the t's.

In general $\mathfrak{M}$ will contain points which are not allowable. But in case the parameters are merely some r of the coördinates themselves it follows readily from a theorem of Ritt (1) that such exceptional points are always limit points of allowable points. It is the purpose of the present paper to extend the above result to all representations, whether the parameters t are essential or not.

THEOREM. *If an algebraic manifold $\mathfrak{M}$ is represented in terms of any parameters whatever, the base field K being the complex field, then each point of $\mathfrak{M}$ is a limit of allowable points.*

We shall treat in detail only the case in which the parameters $t_1, \cdots, t_m$ are all algebraically independent over K; but the method of proof is quite the same if the representation involves additional parameters, say $u_{m+1}, \cdots, u_s$, dependent on the t's. Since each x is algebraic over any extension of $K(t_1, \cdots, t_m)$ we may write the irreducible equation for x_{α_1} over $K(t_1, \cdots, t_m)$, the irreducible equation for x_{α_2} over $K(t_1, \cdots, t_m, x_{\alpha_1})$, etc., each divided through by its leading coefficient:

$$(\alpha)\qquad \begin{array}{l} x^a_{\alpha_1} + a_1(t_1, \cdots, t_m)x^{a-1}_{\alpha_1} + \cdots + a_a(t_1, \cdots, t_m) = 0 \\ x^b_{\alpha_2} + b_1(t_1, \cdots, t_m, x_{\alpha_1})x^{b-1}_{\alpha_2} + \cdots + b_b(t_1, \cdots, t_m, x_{\alpha_1}) = 0 \\ \quad\vdots \\ x^c_{\alpha_n} + c_1(t_1, \cdots, t_m, x_{\alpha_1}, \cdots, x_{\alpha_{n-1}})x^{c-1}_{\alpha_n} + \cdots + c_c(t_1, \cdots, x_{\alpha_{n-1}}) = 0 \end{array}$$

[1] Received November 12, 1936.

the order $x_{a_1}, \cdots, x_{a_n}$ to be determined later. Although the coefficients $a_i, \cdots, c_i$ are in general rational functions of both x's and t's, they may be made polynomials in the x's; the denominators will then involve only $t_1, \cdots, t_m$. Let a non-vanishing common multiple of all denominators be $A(t_1, \cdots, t_m)$. Any set of complex numbers $t'_1, \cdots, t'_m, x'_1, \cdots, x'_n$ satisfying (α) together with the relation $A(t'_1, \cdots, t'_m) \neq 0$ is allowable (2).

The theorem will be proved by showing that for any point $\{x'_1, \cdots, x'_n\}$ of $\mathfrak{M}$ there is a neighboring point $\{\dot{x}_1, \cdots, \dot{x}_n\}$ of to which we can assign parameter values $\dot{t}_1, \cdots, \dot{t}_m$ in such a way that the set $\dot{t}_1, \cdots, \dot{t}_m, \dot{x}_1, \cdots, \dot{x}_n$ satisfies (α) with $A(\dot{t}_1, \cdots, \dot{t}_m) \neq 0$.

For this purpose we choose a new transcendental basis (3) of $K(t_1, \cdots, t_m, x_1, \cdots, x_n)$. Every element of $K(t_1 \cdots, t_m, x_1, \cdots, x_n)$ is algebraically dependent on the ordered set $x_1, \cdots, x_n, t_1, \cdots, t_m$. If from this set we select those elements which are algebraically independent (over K) of all preceding elements, we obtain a set Σ having the following properties:

(a) the number of elements in Σ is exactly m;

(b) every element of $K(t_1, \cdots, t_m, x_1, \cdots, x_n)$ is algebraically dependent on Σ;

(c) the elements of Σ are algebraically independent over K.

Let the elements of Σ be $x_{\beta_1}, \cdots, x_{\beta_r}, t_{\gamma_1}, \cdots, t_{\gamma_{m-r}}$. After suitably renumbering the x's and t's, we may assume that Σ contains $x_1, \cdots, x_r, t_1, \cdots, t_{m-r}$, and that equations (α) have been calculated correspondingly.

The field $K(t_1, \cdots, t_m, x_1, \cdots, x_n)$ can now be regarded as an algebraic extension of the pure transcendental field $K(x_1, \cdots, x_r, t_1, \cdots, t_{m-r})$. We may therefore write the irreducible equation satisfied by x_{r+1} over $K(x_1, \cdots, x_r, t_1, \cdots, t_{m-r})$, the irreducible equation satisfied by x_{r+2} over $K(x_1, \cdots, x_r, t_1, \cdots, t_{m-r}, x_{r+1})$, etc., each divided through by its leading coefficient. Note however that no t's will appear in the coefficients of the equations for $x_{r+1}, \cdots, x_n$; for the existence of a relation $F(x_1, \cdots, x_n, t_1, \cdots, t_{m-r}) = 0$ which actually involved some $t_k \varepsilon \Sigma$ would imply the dependence of that t_k on the set $x_1, \cdots, x_n, t_1, \cdots, t_{k-1}, t_{k+1}, \cdots, t_{m-r}$ and therefore on the set $x_1, \cdots, x_r, t_1, \cdots, t_{k-1}, t_{k+1}, \cdots, t_{m-r}$, which is impossible. The equations under consideration then take the form:

$$(\beta) \quad \begin{aligned} &x^d_{r+1} + d_1(x_1, \cdots, x_r) x^{d-1}_{r+1} + \cdots + d_d(x_1, \cdots, x_r) = 0 \\ &\quad\vdots \\ &x_n^e + e_1(x_1, \cdots, x_r, x_{r+1}, \cdots, x_{n-1}) x_n^{e-1} + \cdots + e_e(x_1, \cdots, x_{n-1}) = 0. \end{aligned}$$

The denominators of the various coefficients need involve only $x_1, \cdots, x_r$; let a non-vanishing common multiple of all denominators be $B(x_1, \cdots, x_r)$.

Continuing, we write the irreducible equation satisfied by t_{m-r+1} over $K(x_1,\cdots,x_r,\ t_1,\cdots,t_{m-r})$, the irreducible equation satisfied by t_{m-r+2} over $K(x_1,\cdots,x_r,\ t_1,\cdots,t_{m-r},t_{m-r+1})$, etc., each divided through by its leading coefficient:

$$(\gamma)\quad \begin{array}{l} t^g_{m-r+1}+g_1(x_i,t_1,\cdots,t_{m-r})\,t^{g-1}_{m-r+1}+\cdots+g_g(x_i,t_1,\cdots,t_{m-r})=0 \\ \quad\vdots \\ t_m{}^h+h_1(x_i,t_1,\cdots,t_{m-r},t_{m-r+1},\cdots,t_{m-1})\,t_m{}^{h-1}+\cdots+h_h(x,t_1,\cdots,t_{m-1})=0. \end{array}$$

Only $x_1,\cdots,x_r,\ t_1,\cdots,t_{m-r}$ need occur in the denominators of the coefficients. Let a non-vanishing common multiple of all denominators be $C(x_1,\cdots,x_r,\ t_1,\cdots,t_{m-r})$.

Any set $x'_1,\cdots,x'_n,\ t'_1,\cdots,t'_m$ of complex numbers satisfying (β) and (γ) with $B\cdot C\neq 0$ will be an allowable set, and will therefore satisfy the equations resulting from (α) on multiplication of each of the latter by $A(t_1,\cdots,t_m)$. It remains then, to ensure the non-vanishing of $A(t'_1,\cdots,t'_m)$. Now as an element of $K(t_1,\cdots,t_m,\ x_1,\cdots,x_n)$, A satisfies some irreducible equation over $K(x_1,\cdots,x_r,\ t_1,\cdots,t_{m-r})$:

$$A^s+\frac{p_1(x_1,\cdots,x_r,t_1,\cdots,t_{m-r})}{q_1(x_1,\cdots,x_r,t_1,\cdots,t_{m-r})}A^{s-1}+\cdots+\frac{p_s(x,t)}{q_s(x,t)}=0.$$

Let $D(x_1,\cdots,x_r,\ t_1,\cdots,t_{m-r})$ be a non-vanishing common multiple of all denominators and of the numerator p_s of the last coefficient. Then $D(x'_1,\cdots,x'_r,\ t'_1,\cdots,t'_{m-r})\neq 0$ implies $A(t'_1,\cdots,t'_m)\neq 0$ for any allowable set $x'_1,\cdots,x'_r,\ t'_1,\cdots,t'_m$. Thus the non-vanishing of the polynomial $B\cdot C\cdot D\equiv V(x_1,\cdots,x_r,\ t_1,\cdots,t_{m-r})$ implies the non-vanishing of all denominators so far considered.

We may arrange V according to power-products of the t's, and take a non-vanishing common multiple $P(x_1,\cdots,x_r)$ of the resulting coefficients. Now P does not vanish everywhere on the irreducible (2) manifold $\mathfrak{M}$, since otherwise it would vanish identically as an element of $K(t_1,\cdots,t_m,x_1,\cdots,x_n)$. Let $\{x'_1,\cdots,x'_n\}$ be any point of $\mathfrak{M}$. If $P(x'_1,\cdots,x'_n)=0$, then by Ritt's theorem that point is a limit point of points $\{\dot{x}_1,\cdots,\dot{x}_n\}$ of $\mathfrak{M}$ for which $P\neq 0$. Suppose therefore that $P(x'_1,\cdots,x'_n)\neq 0$. Now equations (β) after multiplication by $B(x_1,\cdots,x_r)$ are satisfied by all allowable values of the x's, and thus constitute part of the equations defining the manifold $\mathfrak{M}$. But at the point under consideration $B\neq 0$, so that its coördinates actually satisfy equations (β).

The independent quantities $t'_1, \cdots, t'_{m-r}$ may be chosen in such a way that $V(x'_1, \cdots, x'_r, t_1, \cdots, t_{m-r}) \neq 0$. Using these values $x'_1, \cdots, x'_n$, $t'_1, \cdots, t'_{m-r}$, we may calculate successively the remaining t's from (γ). The set $x'_1, \cdots, x'_n, t'_1, \cdots, t'_m$, being allowable, will satisfy equations (α), for by construction $A(t'_1, \cdots, t'_m) \neq 0$.

Thus any point of $\mathfrak{M}$ at which $P \neq 0$ may be obtained directly from the original parametric representation; and any other point of $\mathfrak{M}$ is a limit point of points thus reachable.

PRINCETON UNIVERSITY.

REFERENCES.

(1) J. F. Ritt, "Differential equations," *American Mathematical Society Colloquium Publications*, vol. 14 (1932), p. 91. Also B. L. van der Waerden, "Zur algebraischen Geometrie III," *Mathematische Annalen*, vol. 108 (1933), pp. 694-698.

(2) For notation and definitions cf. B. L. van der Waerden, *Moderne Algebra*, vol. II, pp. 51-61; or B. L. van der Waerden, "Zur Nullstellentheorie der Polynomideale," *Mathematische Annalen*, vol. 96 (1926-27), pp. 183-208.

(3) B. L. van der Waerden, *Moderne Algebra*, vol. I, pp. 204-206.

A REMARK CONCERNING THE PARAMETRIC REPRESENTATION OF AN ALGEBRAIC VARIETY.[1]

By Oscar Zariski.

In his paper "On those points of an algebraic manifold not reachable by a given parametric representation," published in the present issue of this Journal, Mr. J. F. Daly treats the case in which the number of parameters in a given parametric representation of an algebraic variety exceeds the dimension of the variety; nor need the parameters belong to the field of algebraic functions defined by the variety. While this type of parametric representations is more general than the one treated heretofore explicitly in the literature (see, especially, van der Waerden, "Über irreduzible algebraische Mannigfaltigkeiten," *Mathematische Annalen*, vol. 108 (1933)), it may be pointed out that the generalization given by Daly can also be obtained by using some simple properties of rational transformations of varieties. The following proof is taken from the mimeographed notes of the algebraic geometry seminar conducted by Professor Lefschetz and myself in Princeton, 1934.

1. Let V be an irreducible algebraic r-dimensional variety in $S_m(y_1, \cdots, y_m)$ and let

$$x_k = R_k(y_i) = P_k(y_i)/Q(y_i), \qquad (k = 1, 2, \cdots, n) \tag{1}$$

be the equations of a rational transformation of V into an algebraic variety W in $S_n(x_1, \cdots, x_n)$. We assume, of course, that $Q \neq 0$ on V. The coördinates $\eta_1, \eta_2, \cdots, \eta_m$ of a generic point of V are elements of a field $\Omega = K(\eta_1, \cdots, \eta_m)$ of algebraic functions of r independent variables, where K is the field of complex numbers. The coördinates of a generic point of W are $\xi_k = R_k(\eta_i)$ and define a field $\Omega' = K(\xi_k)$, $\Omega' \leqq \Omega$. If ρ $(\leqq r)$ is the degree of transcendentality of Ω', then W is of dimension ρ.

Let $g(y_1, \cdots, y_m)$ be some polynomial which does not vanish identically on V, and let T denote the set of points of W which can be obtained directly from the equations (1) and which correspond to points of V at which $g \neq 0$, i. e. points of W which correspond to points of V at which $Q \neq 0$ and $g \neq 0$. We prove that W *is the closure of* T.

From the theory of fields it follows that $\Omega = \Omega'(t_1, \cdots, t_s, \eta)$, $s + \rho = r$,

[1] Received November 16, 1936.

where the t_j's are algebraically independent over Ω' and where η satisfies an algebraic equation $f(\xi_k, t_l, \eta) = 0$ with coefficients in K. Let

$$\eta_i = S_i(\xi_k, t_l, \eta)/M(\xi_k, t_l),$$

where S_i and M are polynomials, and let

$$N(Q) = L(\xi_k, t_l)/M(\xi_k, t_l), \qquad N(g) = G(\xi_k, t_l)/M(\xi_k, t_l)$$

be the norms over $K(\xi_k, t_l)$ of $Q(\eta_1, \cdots, \eta_m)$ and of $g(\eta_1, \cdots, \eta_m)$ respectively. Here $M(\xi_k, t_l) \neq 0$, and also $L(\xi_k, t_l) \neq 0$, $G(\xi_k, t_l) \neq 0$, since $Q(\eta_i) \neq 0$ and $g(\eta_i) \neq 0$, by hypothesis. Let then x^0 be a point of W at which the polynomials in $t: M(x^0{}_k, t_l)$, $L(x^0{}_k, t_l)$, $G(x^0{}_k, t_l)$ do not vanish identically. If $(t^0{}_1, \cdots, t^0{}_s)$ is any set of values of the t's at which these polynomials do not vanish, and if η^0 is a root of $f(x^0{}_k, t^0{}_l, \eta) = 0$, then $y^0{}_i = S_i(x^0{}_k, t^0{}_l, \eta^0)/M(x^0{}_k, t^0{}_l)$ are the coördinates of a point of V at which $Q \neq 0$, $g \neq 0$, and moreover $x^0{}_k = R_k(y^0{}_i)$. That is, any point (x) of W at which none of the polynomials $M(x, t)$, $L(x, t)$, $G(x, t)$ vanishes identically belongs to the set T. It follows then by a theorem of Ritt that every point of W is a limit point of points in T, q. e. d.

2. Let the coördinates $x_1, \cdots, x_n$ of a generic point of an algebraic ρ-dimensional variety W be algebraic functions of r parameters $t_1, \cdots, t_r$, independent over K. The variety W is a rational transform of the r-dimensional variety whose generic point is $(x_1, x_2, \cdots, x_n, t_1, \cdots, t_r)$. We identify this variety with the variety V of the preceding section. For the variety V the parameters t_l are merely some of the coördinates, and the points of V which cannot be reached by this parametric representation satisfy a certain equation $g = 0$, where $g = g(x, t)$ is a polynomial not identically zero on V. By the preceding section it follows that the points of W which cannot be reached by the given parametric representation are limit points of reachable points.

THE JOHNS HOPKINS UNIVERSITY.

ON THE CONSTRUCTION OF SYMMETRIC RULED SURFACES.[1]

By Arnold Emch.

1. *Introduction.* Let $S_3(x)$ denote a projective space of three dimensions with the homogeneous variables x_1, x_2, x_3, x_4 and $\phi_1, \phi_2, \phi_3, \phi_4$ the four elementary symmetric functions on these variables, then

$$(1) \qquad \Sigma a_k \phi_1^{n_1} \phi_2^{n_2} \phi_3^{n_3} \phi_4^{n_4} = 0,$$

with the n_i as positive integers, and $n_1 + 2n_2 + 3n_3 + 4n_4 = n$ satisfied represents, by definition, a symmetric surface. Noted examples are the Cayley cubic $\Sigma 1/x_i = 0$ and Clebsch's diagonal surface. The writer has investigated surfaces of this kind in a number of papers. Two symmetric surfaces intersect obviously in a symmetric space curve, which is left invariant by the symmetric group G_{24} of collineations, since the surfaces producing it are invariant. In this connection I mention the discovery of a remarkable sextic of genus four which lies on 10 cubic cones.[2]

It is naturally interesting to enquire about the possibility of symmetric ruled surfaces. It is clear that such surfaces exist, because $\phi_1^2 + \lambda\phi_2 = 0$ represents a pencil of symmetric quadrics (admitting imaginary rulings). Then, of course, there is the special class of symmetric cones. If we admit the existence of general symmetric ruled surfaces, then by G_{24} a generic generatrix determines immediately 23 more on the surface. The first question then is what is the lowest order of surface on 24 lines belonging to the group G_{24}. Choose a generic line l, determined parametrically by

$$(2) \qquad \rho x_i = a_i + \lambda b_i \qquad (i = 1, 2, 3, 4),$$

with (a) and (b) as two arbitrary distinct points. Substituting (2) in (1), l will lie on (1) when (2) is identically equal to zero for all values of λ. This leads to an equation of degree n in λ, which is satisfied for all values of λ when its $n + 1$ coefficients vanish. Hence the number of coefficients in (1) must have a number of coefficients (effective) $\geqq n + 1$. The experimental solution of the Diophantine equation gives for the orders $n = 1, 2, 3, 4, 5, 6, 7, \cdots$ the number of effective constants of (1): $0, 1, 2, 3, 4, 7, 9, \cdots$. This shows

[1] Received October 21, 1936.

[2] "Über eine besondere Raumkurve sechster Ordnung," *Monatshefte für Mathematik und Physik*, vol. 40 (1933), pp. 193-200.

that the first case in which the inequality is satisfied is for $n = 6$. There is still one constant available. Hence

THEOREM 1. *On a generic line l in S_3 there is a pencil of symmetric sextic surfaces. There are no such surfaces of lower order on l.*

For general ruled surfaces one must look for higher than the 6th order. To solve this problem we use the parametric representation.

2. *Construction of ruled symmetric surfaces.* Consider the equations

$$
\begin{aligned}
\rho x_1 &= \phi^*{}_1(\lambda_2, \lambda_3, \lambda_4) + \mu\psi^*{}_1(\lambda_2, \lambda_3, \lambda_4),\\
\rho x_2 &= \phi^*{}_2(\lambda_1, \lambda_3, \lambda_4) + \mu\psi^*{}_2(\lambda_1, \lambda_3, \lambda_4),\\
\rho x_3 &= \phi^*{}_3(\lambda_1, \lambda_2, \lambda_4) + \mu\psi^*{}_3(\lambda_1, \lambda_2, \lambda_4),\\
\rho x_4 &= \phi^*{}_4(\lambda_1, \lambda_2, \lambda_3) + \mu\psi^*{}_4(\lambda_1, \lambda_2, \lambda_3),
\end{aligned}
\tag{3}
$$

in which $\phi^*(\alpha, \beta, \gamma)$ and $\psi^*(\alpha, \beta, \gamma)$ are symmetric polynomials in α, β, γ, from which the $\phi^*{}_i$ and $\psi^*{}_i$ are obtained by replacing α, β, γ by $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ as indicated. Then the $\phi^*{}_i$ and $\psi^*{}_i$, except as to a possible change of signs throughout, permute in the same way as any chosen permutations of the λ's. Hence a permutation

$$
\begin{pmatrix} \lambda_1 & \lambda_2 & \lambda_3 & \lambda_4 \\ \lambda_i & \lambda_k & \lambda_l & \lambda_g \end{pmatrix} \text{ induces the permutation } \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \\ \rho x_i & \rho x_k & \rho x_l & \rho x_g \end{pmatrix},
$$

in which $\rho = \pm 1$. For a definite set of values for the λ's the ϕ^*'s and ψ^*'s in (3) represent two points in S_3, and with μ variable (3) represents a straight line l. Thus

THEOREM 2. *On the application of G_{24} to a definite set $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ and with μ variable (3) represents 24 lines l_i. Each of these is met by six others of the same set.*

To prove the second part of the theorem, notice that a line l_i cuts the six planes $x_i - x_k = 0$ in six points P_{ik}. By the substitution (ik) this P_{ik} is not changed, but l_i is transformed into that of set of 24 which corrresponds to (ik).

To (3) we now adjoin

$$
\begin{aligned}
M(\lambda_1, \lambda_2, \lambda_3, \lambda_4) &= 0,\\
N(\lambda_1, \lambda_2, \lambda_3, \lambda_4) &= 0,
\end{aligned}
\tag{4}
$$

two equations of degree m and n and symmetric in the λ's, and also homogeneous in order to make them of dimension 3. Then in a $S_3(\lambda)$-space (4)

represents two surfaces which we assume to intersect in a complete irreducible curve C_{mn}. To every point (λ) of this curve correspond by the G_{24} 24 points (including the point itself) of the same curve and by (3) 24 lines l_i. As (λ) describes C_{mn}, the l_i's generate a ruled surface R whose genus is the same as that of C_{mn}, because there exists a $(1,1)$ correspondence between the points of C_{mn} and the generatrices of the ruled surface in $S_3(x)$. To determine the order of R, the λ's and μ must be eliminated from (3) and (4). Denote $\phi^*_i(\lambda_j, \lambda_k, \lambda_l)$ by ϕ^*_i, then elimination of μ from (3) gives

$$(5)\quad \begin{aligned} &(a)\ x_1(\phi^*_3\psi^*_4 - \phi^*_4\psi^*_3) + x_3(\phi^*_4\psi^*_1 - \phi^*_1\psi^*_4) + x_4(\phi^*_1\psi^*_3 - \phi^*_3\psi^*_1) = 0,\\ &(b)\ x_2(\phi^*_3\psi^*_4 - \phi^*_4\psi^*_3) + x_3(\phi^*_4\psi^*_2 - \phi^*_2\psi^*_4) + x_4(\phi^*_2\psi^*_3 - \phi^*_3\psi^*_2) = 0.\end{aligned}$$

These are equations of degree $r+s$ in the λ's. Elimination of the λ's gives a symmetric equation of degree $mn(r+s)$ in x_1, x_3, x_4 and of degree[3] $mn(r+s)$ in x_2, x_3, x_4, hence of degree $mn(r+s)$ in x_1, x_2, x_3, x_4. If however equations (4) and (5) have k points in common which are in the same order $\alpha, \beta, \gamma, \delta$-fold four the four equations, then the resultant is of degree $n(r+s)^2 - k\beta\gamma\delta$ in the coefficients of M; $n(r+s)^2 - k\alpha\gamma\delta$ in the coefficients of N; $mn(r+s) - k\alpha\beta\delta$ in the coefficients of (5_a); $mn(r+s) - k\alpha\beta\gamma$ in the coefficients of (5_b). When $\delta = \gamma$, then the resultant (reduced) is of degree $mn(r+s) - k\alpha\beta\gamma$. Hence

THEOREM 3. *The parametric equations* (3) *in conjunction with* (4) *and the indicated multiplicities represent a ruled symmetric surface R of order $mn(r+s) - k\alpha\beta\gamma$.*

3. *Example of octic ruled surface.* If (3) has the form

$$(6)\quad \begin{aligned} \rho x_1 &= \lambda_2\lambda_3\lambda_4 + \mu\lambda_1,\\ \rho x_2 &= \lambda_1\lambda_3\lambda_4 + \mu\lambda_2,\\ \rho x_3 &= \lambda_1\lambda_2\lambda_4 + \mu\lambda_3,\\ \rho x_4 &= \lambda_1\lambda_2\lambda_3 + \mu\lambda_4,\end{aligned}$$

and

$$(7)\quad \begin{aligned} M &= \Sigma\lambda_i\lambda_k,\\ N &= \Sigma\lambda_i\lambda_k\lambda_l;\end{aligned}$$

$k = 4$ (vertices of coördinate tetrahedrons), $\alpha = 1$, $\beta = 2$, $\delta = \gamma = 2$, then

[3] The resultant of equations (4) and (5) according to the conventional methods is of degree $2mn(r+s)$, but reduces to the degree $mn(r+s)$ after deleting extraneous factors. This is verified by means of the $(1,1)$ correspondence which exists between the curves C_{mnr} and C_{mns} of indicated orders, obtained in a λ-space by the mapping of C_{mn} $(M \times N)$ by means of the ϕ^*'s and ψ^*'s.

the order of R becomes $3 \cdot 2 \cdot 4 - 4 \cdot 2 \cdot 2 = 8$. Ordinarily the elimination of the λ's and μ presents great difficulties on account of the labor involved. In this example this task may actually be accomplished as follows: Let $\phi_1, \phi_2, \phi_3, \phi_4$ now stand for the elementary symmetric functions of the variables x, and L_1, L_2, L_3, L_4 for the λ's. Then

$$
(8)\quad
\begin{aligned}
\rho\, \phi_1 &= L_3 + \mu L_1 \\
\rho^2\phi_2 &= L_2L_4 + \mu(L_1L_3 - 4L_4) + \mu^2 L_2 \\
\rho^3\phi_3 &= L_1L_4^2 + \mu L_4(L_1L_2 - L_3L_4) + \mu^2(L_2L_3 - 3L_1L_4) + \mu^3 L_3 \\
\rho^4\phi_4 &= L_4^3 + \mu L_4^2(L_1^2 - 2L_2) \\
&\quad + \mu^2 L_4(L_2^2 - 2L_1L_3 + 2L_4) + \mu^3(L_3^2 - 2L_2L_1) + \mu^4 L_4.
\end{aligned}
$$

When (λ) lies on the intersection C_6 (rational with double points at the A_i) of $L_2 = 0$, $L_3 = 0$, (8) reduces to

$$
(9)\quad
\begin{aligned}
\rho\, \phi_1 &= \mu L_1, \\
\rho^2\phi_2 &= -4\mu L_4, \\
\rho^3\phi_3 &= L_1L_4^2 - 3\mu^2 L_1L_4, \\
\rho^4\phi_4 &= L_4^3 + \mu L_1^2L_4^2 + 2\mu^2 L_4^2 + \mu^4 L_4.
\end{aligned}
$$

Eliminating ρ, μ, L_1, L_4 from (9) we obtain the octic ruled surface R:

$$
(10)\quad
\begin{aligned}
16\phi_1\phi_3\phi_4 - 8\phi_1\phi_2^2\phi_3 - 12\phi_1^2\phi_2\phi_4 + 4\phi_1^2\phi_2^3 - 16\phi_1^2\phi_3^2 \\
+ 24\phi_1^3\phi_2\phi_3 - 9\phi_1^4\phi_2^2 + 4\phi_2\phi_3^2 = 0.
\end{aligned}
$$

It cuts the unit plane $\phi_1 = 0$ in the conic $(\phi_1 = 0, \phi_2 = 0)$ C_2 and contains the three diagonal lines $(\phi_1 = 0, \phi_3 = 0)$ as double lines. If $P(\lambda)$ lies on C_6, then by the cubic involution $\rho\lambda'_i = 1/\lambda_i$, $P(\lambda)$ goes into $P'(1/\lambda_1, 1/\lambda_2, 1/\lambda_3, 1/\lambda_4)$ which lies in the intersection of L_1 and L_2, i. e., on C_2. The join of $P'P$ interpreted in S_3 is a generatrix of R. Thus

THEOREM 4. *The locus of joins of corresponding points in the cubic involution T in which C_2 and C_6 are corresponding is an octic symmetric ruled surface R.*

This can be verified by the principle of correspondence: Let g be a generic line in S_3 and (α) the pencil of planes on g. An α cuts C_6 in six points B: to which correspond by T six points B'_i on C_2, which joined to g give six planes α'. Conversely to a plane α' which cuts C_2 in two points B'_i correspond by T two points B_i on C_6, thus determining two planes α. This establishes a $(6, 2)$-correspondence between the planes α and α' with $6 + 2 = 8$ coincidences. Let α^* be such a coincidence, then there lie on this plane two corresponding

points by T, P on C_6 and P' on P_2, such that $P'P$ is a generatrix of R. There are thus 8 such generatrices cutting g. R is of order 8.

4. *Developable ruled surfaces.* Every developable surface which is not a cone has an "edge of regression" or cuspidal curve which may be any space curve. When the developable surface is symmetric, then a tangent of the cuspidal curve is transformed into 24 other such tangents by the G_{24}. From this follows that the cuspidal curve is symmetric and that it may therefore be obtained as the intersection of two symmetric surfaces $f_m = 0$ and $g_n = 0$ of orders m and n respectively. I shall restrict myself to the case in which the intersection of f_m and g_n is a complete irreducible curve C_{mn}. The tangent planes at a point of intersection (y) of the two surfaces, to each f_m and g_n are

$$(11) \qquad \Sigma x_i \frac{\partial f}{\partial y_i} = 0, \qquad\qquad (12) \qquad \Sigma x_i \frac{\partial g}{\partial y_i} = 0.$$

They intersect in a tangent t of the curve of intersection of the two surfaces and as (y) describes the curve of intersections C_{mn}; t describes the developable surface D which is now symmetric. Its order is obtained by eliminating (y) from $f_m = 0$, $g_n = 0$, (11) and (12), which gives for the order of D $mn(n-1) + mn(m-1)$ or

$$(13) \qquad d = mn(m+n-2).$$

The order of the double curve of D, since C_{mn} has supposedly no effective singularities is

$$(14) \qquad \vartheta = \tfrac{1}{2}mn(m+n-2)[mn(m+n-2)-4],$$

according to a well known formula (Cayley). D has a double curve of order $\frac{1}{2}mn(m+n-2)$ (which is always possible, because d is even) in each of the six planes $x_i - x_k = 0$. Hence, outside of these components, the double curve of D contains a residual double curve of order

$$(15) \qquad \sigma = \tfrac{1}{2}mn(m+n-2)[mn(m+n-2)-10].$$

To find the point of intersection of t with the unit plane, we solve (11), (12) and $\Sigma x_i = 0$ for (x). Denoting $\partial f/\partial y_i$ and $\partial g/\partial y_i$ by f_i and g_i respectively

$$(16) \qquad \begin{aligned} \rho x_1 &= f_3g_4 - f_4g_3 + f_4g_2 - f_2g_4 + f_2g_3 - f_3g_2 \\ \rho x_2 &= -(f_3g_4 - f_4g_3 + f_4g_1 - f_1g_4 + f_1g_3 - f_3g_1) \\ \rho x_3 &= f_2g_4 - f_4g_2 + f_4g_1 - f_1g_4 + f_1g_2 - f_2g_1 \\ \rho x_4 &= -(f_2g_3 - f_3g_2 + f_3g_1 - f_1g_3 + f_1g_2 - f_2g_1). \end{aligned}$$

As the point $P(\lambda)$ describes C_{mn}, the point (x) in (16) describes the curve of intersection D' of D with the unit plane. The join of D' with P is a tangent of C_{mn} or a generatrix of D. With μ variable we have

THEOREM 5. *The parametric representation of the developable tangent surface of C_{mn} the curve of intersection of f_m and g_n is given by*

$$
\begin{aligned}
\sigma x_1 &= \quad f_3g_4 - f_4g_3 + f_4g_2 - f_2g_4 + f_2g_3 - f_3g_2 + \mu\lambda_1 \\
\sigma x_2 &= -(f_3g_4 - f_4g_3 + f_4g_1 - f_1g_4 + f_1g_3 - f_3g_1) + \mu\lambda_2 \\
\sigma x_3 &= \quad f_2g_4 - f_4g_2 + f_4g_1 - f_1g_4 + f_1g_2 - f_2g_1 + \mu\lambda_3 \\
\sigma x_4 &= -(f_2g_3 - f_3g_2 + f_3g_1 - f_1g_3 + f_1g_2 - f_2g_1) + \mu\lambda_4; \\
& f_m = 0,\ g_n = 0.
\end{aligned}
\tag{17}
$$

As may be expected these are precisely of the type represented by (3) and (4).

Example. The simplest developable symmetric surface is obtained as the tangent surface of the sextic C_6 of genus four obtained as the intersection of the symmetric cubic and quadric

$$M = \Sigma\lambda_i\lambda_k\lambda_l = 0 \quad\text{and}\quad N = \Sigma\lambda_i\lambda_k, \qquad (i \neq k \neq l = 1, 2, 3, 4).$$

In this case (17) becomes

$$
\begin{aligned}
\rho x_1 &= -(\lambda_2 - \lambda_3)(\lambda_3 - \lambda_4)(\lambda_4 - \lambda_2) + \mu\lambda_1 \\
\rho x_2 &= \quad (\lambda_1 - \lambda_3)(\lambda_3 - \lambda_4)(\lambda_4 - \lambda_1) + \mu\lambda_2 \\
\rho x_3 &= -(\lambda_1 - \lambda_2)(\lambda_2 - \lambda_4)(\lambda_4 - \lambda_1) + \mu\lambda_3 \\
\rho x_4 &= \quad (\lambda_1 - \lambda_2)(\lambda_2 - \lambda_3)(\lambda_3 - \lambda_1) + \mu\lambda_4
\end{aligned}
\tag{18}
$$

a symmetric developable surface of order 12, whose equation in the x's is obtained by elimination of ρ, μ, and the λ's from (18) and $M = 0$, $N = 0$. This is omitted on account of the enormous amount of labor which is required.

UNIVERSITY OF ILLINOIS.

ON CIRCLES CONNECTED WITH THREE AND FOUR LINES.*

By J. R. MUSSELMAN.

I. Let us consider six points a_i $(i = 1, 2 \cdots 6)$ of a plane (or sphere) which are ordered. A quadratic covariant associated with them has been discussed by the Morleys.[1] Here we study the six points under the condition that the cross-ratio

$$(1.1) \qquad \frac{(a_1 - a_2)(a_3 - a_4)(a_5 - a_6)}{(a_2 - a_3)(a_4 - a_5)(a_6 - a_1)} = \rho,$$

where ρ is any real number. Since this cross-ratio is invariant under homographies

$$y = (\alpha x + \beta)/(\gamma x + \delta)$$

we are asking that it be invariant also under antigraphies

$$\bar{y} = (\alpha x + \beta)/(\gamma x + \delta).$$

The significance of (1.1) is easily seen. For the equation

$$\frac{(a_1 - a_2)(a_3 - x)}{(a_2 - a_3)(x - a_1)} = \rho_1$$

represents a circle on the points a_1, a_2 and a_3. Writing similar expressions for the circles $a_3a_4a_5$ and $a_5a_6a_1$, and multiplying the three equations together we see that (1.1) is precisely the condition *that circles* $a_1a_2a_3$, $a_3a_4a_5$, $a_5a_6a_1$ *have a common point*. Furthermore, the condition that circles $a_2a_3a_4$, $a_4a_5a_6$, $a_6a_1a_2$ have a common point is

$$(1.2) \qquad \frac{(a_2 - a_3)(a_4 - a_5)(a_6 - a_1)}{(a_3 - a_4)(a_5 - a_6)(a_1 - a_2)} = \rho'.$$

Now the truth of either one of (1.1) or (1.2) implies the truth of the other, hence we have here a simple proof of the theorem *if the circles* $a_1a_2a_3$, $a_3a_4a_5$, $a_5a_6a_1$ *meet at a point, say* m; *then the circles* $a_2a_3a_4$, $a_4a_5a_6$, $a_6a_1a_2$ *meet at a point, say* n.

If we choose the points a_i so that m is ∞, then a_2, a_4, a_6 lie respectively on the lines a_1a_3, a_3a_5 and a_5a_1. This gives us then the theorem of Miquel—

* Read before the International Congress of Mathematicians, Oslo, July 17, 1936. Received by the Editors, November 30, 1936; revised February 4, 1937.

[1] *Inversive Geometry*, p. 60. See also exercise 6, page 30.

if a point be marked on each side of a triangle, and through each vertex of the triangle and the marked points on the adjacent sides a circle be drawn, the three circles meet at a point. We notice also that the homography which sends a_1 into a_4, a_3 into a_6, and a_5 into a_2 sends the line $a_3a_5\infty$ into the circle a_6a_2n, so that in general a_1a_4, a_3a_6, a_5a_2, and mn are pairs in a homography.

II. In this section we connect the above with the following generalization of an old theorem[2] that *if a point p has images a_2, a_4, a_6 in the sides of a triangle $a_1a_3a_5$, then the circles $a_2a_3a_4$, $a_4a_5a_6$, $a_6a_1a_2$ meet at a point, say n, on the circle $a_1a_3a_5$: and equally the circles $a_1a_2a_3$, $a_3a_4a_5$, $a_5a_6a_1$ meet at a point, say m, on the circle $a_2a_4a_6$.* We give an analytic proof of the first part which enables us to construct the points m and n without the aid of the circles. Let the coördinates of the vertices of the triangle $a_1a_3a_5$ be turns t_i, i. e. $|t_i| = 1$. The elementary symmetric functions of the t_i will be denoted by σ_i. The reflection a_2 of the point p in the side a_1a_3 will have as coördinate $t_1 + t_3 - t_1t_3\bar{p}$. Consider the equation

$$(t_1 + t_3)x = \sigma_2 - \sigma_3\bar{p} + (t_1t_3 - \sigma_3\bar{p})T. \tag{2.1}$$

For $T = -1$, $t_3t_1^{-1}$, $t_1t_3^{-1}$ respectively, this circle (2.1) passes through the points a_5, a_4, a_6. Also for the turn value

$$T = -\frac{(\sigma_2 - \sigma_3\bar{p})(t_5 - p)}{(\sigma_1 - p)(t_1t_3 - \sigma_3\bar{p})}, \qquad x = \frac{\sigma_2 - \sigma_3\bar{p}}{\sigma_1 - p},$$

a point on the circle $a_1a_3a_5$. Since the coördinate of this point, n, is symmetric in t_1, t_3, t_5, the circles $a_6a_1a_2$ and $a_2a_3a_4$ likewise are on n, which proves the first part of the theorem. The second part follows immediately from the homography which connects the pairs a_1a_4, a_3a_6, a_5a_2, m and n, for it is of period two.

For if p and a_2 are images in the side a_1a_3 then $(a_1 - a_2)/(a_2 - a_3)$ and $(a_1 - p)/(p - a_3)$ must be conjugates. Writing similar expressions for p and a_4, p and a_6 and multiplying the three we obtain

$$\frac{(a_1 - a_2)(a_3 - a_4)(a_5 - a_6)}{(a_2 - a_3)(a_4 - a_5)(a_6 - a_1)} = 1.$$

But this is the condition that a_1a_4, a_2a_5, a_3a_6 be pairs in an involution. Naturally m and n belong to this involution. When m is ∞, we have the theorem of Menelaus.

[2] This is a generalization of a theorem of Canon, *Nouvelles Annales de Mathématique*, Fourth Series, vol. 8 (1908), p. 480, Problem 2108. See also R. Bouvaist, *ibid.*, vol. 10 (1910), p. 136.

Now if H be the orthocenter of $a_1a_3a_5$, any point on the line Hp will have as coördinate $r = (\sigma_1 + \lambda p)/(1 + \lambda)$. The coördinate of n will be $(\sigma_2 - \sigma_3\bar{r})/(\sigma_1 - r)$. Substituting the values for r and $\bar{r}$ in this expression for n we obtain $(\sigma_2 - \sigma_3\bar{p})/(\sigma_1 - p)$. Since this is independent of the parameter λ, we see that *the point n is a fixed point on $a_1a_3a_5$ for all points on the line Hp.* The Simson line of n is parallel to Hp, so n can be constructed[3] without the use of circles. The coördinate of the point m can be written as

(2.2) $$m = \sigma_1 - p + \frac{p - n}{1 - \bar{p}n}, \qquad n = \frac{\sigma_2 - \sigma_3\bar{p}}{\sigma_1 - p}.$$

The line pn cuts the circle $a_1a_3a_5$ at a point L, whose coördinate is

(2.3) $$L = \frac{p - n}{1 - \bar{p}n}.$$

Hence if O be the center of the circle $a_1a_3a_5$, the point n can be constructed from the vector relation

(2.4) $$\overline{Om} = \overline{OH} - \overline{Op} + \overline{OL}.$$

III. Casey[4] defines as "twin points" any two points such that the angles subtended by the sides of a triangle at these points are either equal or supplementary. With reference to the triangle $a_1a_3a_5$ the points p and m are twin points. It is also known that any two points at the ends of a diameter of a rectangular hyperbola are twin points with reference to any triangle inscribed in the hyperbola. The theorem in Section II connecting p and m enables us to state two interesting theorems about the rectangular hyperbola. (1) *If we reflect any point p of a rectangular hyperbola in a chord a_1a_3, obtaining the point a_2, then the circle $a_1a_2a_3$ will intersect the hyperbola at m, the diametrically opposite point of p.* (2) *Let A_1, A_2, A_3, P and Q be any five points of a rectangular hyperbola with P and Q at the ends of a diameter, and let P_i, Q_i $(i = 1, 2, 3)$ be the reflections of P and Q respectively in the chords A_2A_3, A_3A_1, A_1A_2; then the circles $A_2A_3P_1$, $A_3A_1P_2$, $A_1A_2P_3$ and $P_1P_2P_3$ will meet at Q and the circles $A_2A_3Q_1$, $A_3A_1Q_2$, $A_1A_2Q_3$ and $Q_1Q_2Q_3$ will meet at P.*

One application to the geometry of the triangle will indicate the importance of this theorem. The isogenic centers or Fermat points F, F' of a triangle $A_1A_2A_3$ lie on the ends of a diameter of the Kiepert hyperbola which also passes through the vertices of the triangle. Hence *if we reflect either Fermat point in the sides A_2A_3, A_3A_1, A_1A_2, obtaining the points B_1, B_2 and B_3 the circles $A_2A_3B_1$, $A_3A_1B_2$, $A_1A_2B_3$, $B_1B_2B_3$ are on the other Fermat point.*

[3] See R. A. Johnson, *Modern Geometry*, page 208, Theorem 329.

[4] *A Sequel to Euclid*, Sixth Edition, page 249.

IV. We consider here the reflections of a point p in four lines. Our coördinate system is chosen so that the parabola which touches the four lines has the form

$$x = \frac{2}{(1-t)^2}.$$

If p be the focus of the parabola, and we reflect it in the triangle formed by the tangents at the points t_1, t_2 and t_3 the coördinate of the point n will be

$$n = \frac{2}{1-\sigma_1} = \frac{2}{1-S_1+t_4},$$

where S_1 is the symmetric function of four points t_i. For each triangle formed by three out of the four lines we shall have a point n, and these four points n lie on the circle C_4, whose equation is

$$x = \frac{2}{1-S_1+t}.$$

The inverse of the focus of the parabola as to this circle is the point

$$z = \frac{2}{1-S_1}.$$

For five lines of a parabola, we can construct five circles as C_4, and the inverse point of the focus in each circle will lie on a circle, namely on

$$x = \frac{2}{1-\sigma_1+t},$$

where now σ_1 refers to five points t_1. The chain runs on indefinitely for lines of a parabola.

When p is a point on the circumcircle of the triangle, we see from (2. 3) and (2. 4) that L coincides with p and that m is at the orthocenter. Hence *if we reflect the focus of a parabola touching any four lines in the four lines, the four points m, one for each three of the four lines, lie on a line—the directrix of the parabola.*

For p any point in the plane, the coördinate of n is

$$n = \frac{2(p+\bar{p}-2)}{\Pi p + 2(\sigma_1 - 1)}, \qquad (\Pi = 1-\sigma_1+\sigma_2-\sigma_3).$$

The numerator of this fraction equated to zero is the equation of the directrix; the denominator equated to zero gives the orthocenter of the triangle of tangents at the points t_1, t_2, t_3. We thus have the theorem *if we reflect any point*

on the directrix of a parabola in the triangle formed by three of its tangents, the point n coincides with the focus. Naturally n is indeterminate when p is the orthocenter of the triangle of tangents.

V. If we choose as the point p the circumcenter of a triangle $A_1A_2A_3$, then from (2. 2) the coördinate of m can be written as

$$m_4 = \sigma_1 - \sigma_2/\sigma_1$$

where σ_i are symmetric functions of the three symbols t_i. If we take a fourth point A_4 on the circle $A_1A_2A_3$ we can form four triangles from the four points, each determining a point m_i, which four points m_i lie on a circle. For

$$m_4 = \sigma_1 - \sigma_2/\sigma_1 = \frac{S_1{}^2 - S_2 - t_4S_1}{S_1 - t_4},$$

where S_i are symmetric functions of the four t_i. This, for variable t_4, is the equation of a circle; hence the four points m_i lie on the circle C_5, whose equation may be written as

$$x = \frac{a_2 - a_1t}{a_1 - t}$$

where we write a_2 for $S_1{}^2 - S_2$ and a_1 for S_1. Now when $t = 0$, $x = a_2/a_1$; when $t = \infty$, $x = a_1$. Therefore, the points a_2/a_1 and a_1 are inverse points as to this circle. Four other pairs of points in the involution set up by this circle are A_4m_4, A_3m_3, A_2m_2, A_1m_1. Hence *the five pairs of points a_2/a_1 and a_1, A_i and m_i are pairs of an involution.* Since

$$\frac{a_2}{a_1} = \frac{a_2 - a_1t_5}{a_1 - t_5}$$

where a_i are the functions S_1 and $S_1{}^2 - S_2$ written for five symbols t_i, we have shown that if we take five points on a circle, we can determine for each four of the five points a circle such as C_5 and a point such as a_1, *and the inverse point of each a_1 as to its associated circle gives five points which lie on a circle.* This chain can go on indefinitely.

WESTERN RESERVE UNIVERSITY.

ON THE DENSITIES OF INFINITE CONVOLUTIONS.[1]

By Aurel Wintner.

It has been pointed out in a previous paper [2] that, due to certain compactness properties of uniformly bounded functions of uniformly bounded variation, there exists for the "term-by-term" differentiation of infinite convolutions of distribution functions a theorem which does not assume any convergence property of the sequence of derivatives and is, therefore, useful [3] in applications. The present note proves an essential refinement of the theorem in question by showing that, for the "smooth" term of the infinite convolution, the existence of an additional derivative need not be required.[4] In fact, it will be shown that *if at least one term σ_j of a convergent infinite convolution $\sigma_1 * \sigma_2 * \cdots$, say the term $\sigma_j = \sigma_1$, has for $-\infty < x < +\infty$ a continuous density of bounded variation, then so does the distribution function σ represented by the infinite convolution; and the continuous density of $\sigma_1 * \cdots * \sigma_n$ tends, as $n \to +\infty$, to that of $\sigma = \sigma_1 * \sigma_2 * \cdots$ uniformly in every fixed bounded x-range.*

Let $V(\psi)$ denote the total variation $(\leqq +\infty)$ of a function $\psi(x)$, where $-\infty < x < +\infty$. Suppose that the derivatives $\phi'_n(x)$ of a sequence of uniformly bounded, differentiable functions $\phi_n(x)$, $-\infty < x < +\infty$, are such that

(i) $\{\phi'_n(x)\}$ is equicontinuous for $-\infty < x < +\infty$;
(ii) $|\phi'_n(x)| \leqq C$ for some constant C;
(iii) $V(\phi'_n) \leqq c$ for some constant c.

Then [5] the sequence $\{\phi_n(x)\}$ cannot tend to a limit function $\lim \phi_n(x)$ almost everywhere unless $\lim \phi_n(x)$ determines for $-\infty < x < +\infty$ a continuous function which has a uniformly continuous derivative of bounded variation, in which case

$$\phi_n(x) \to \lim_{n\to\infty} \phi_n(x) \quad \text{and} \quad \phi'_n(x) \to (\lim_{n\to\infty} \phi_n(x))'$$

[1] Received January 28, 1937.

[2] A. Wintner, *American Journal of Mathematics*, vol. 57 (1935), pp. 363-366.

[3] Cf. E. R. van Kampen and A. Wintner, *ibid.*, vol. 59 (1937), p. 186 and p. 203.

[4] The assumption made *loc. cit.*[2] was that σ_1 has for $-\infty < x < +\infty$ an absolutely integrable and bounded second derivative. It is clear that this assumption implies the existence of a continuous first derivative which has for $-\infty < x < +\infty$ a bounded variation. Incidentally, a function of bounded variation, which is a derivative, cannot have discontinuities.

[5] This is a corollary of the facts proved *loc. cit.*,[2] pp. 364-365.

hold uniformly in every fixed bounded x-range. Now if $\phi_n = \sigma_1 * \cdots * \sigma_n$, then $\{\phi_n(x)\}$ is uniformly bounded, since ϕ_n is then a distribution function. Furthermore, $\phi_n(x) \to \sigma(x)$ holds almost everywhere, since $\sigma = \sigma_1 * \sigma_2 * \cdots$ is supposed to be a convergent infinite convolution. Hence it is sufficient to prove that the assumption of a derivative $\sigma'_1(x)$, $-\infty < x < +\infty$, of bounded variation implies for the finite convolutions $\phi_n = \sigma_1 * \cdots * \sigma_n$ the existence of densities $\phi'_n = \phi'_n(x)$ which satisfy (i), (ii) and (iii), no matter what the distribution functions $\sigma_2, \sigma_3, \cdots$ may be.

First, if $\omega = \omega(\delta)$ denotes, for a fixed $\delta > 0$, the greatest lower bound of those numbers $\beta > 0$ which have the property that $|\sigma'_1(x^1) - \sigma'_1(x^2)| \leqq \beta$ is a consequence of $|x^1 - x^2| \leqq \delta$ for arbitrary x^1, x^2, then $\omega(\delta) \to 0$ as $\delta \to 0$. In other words, $\sigma'_1(x)$ is not only continuous but uniformly continuous for $-\infty < x < +\infty$. In fact, $\sigma'_1(-\infty) = 0$ and $\sigma'_1(+\infty) = 0$, since the distribution function $\sigma_1(x)$ is supposed to be such that $V(\sigma'_1) < +\infty$. This also implies that $0 \leqq \sigma'_1(x) < V(\sigma'_1)$ for every x.

On placing $\tau_n = \sigma_2 * \cdots * \sigma_n$, it is clear that

$$\phi_n(x) = \phi_n = \sigma_1 * \cdots * \sigma_n = \sigma_1 * \tau_n = \int_{-\infty}^{+\infty} \sigma_1(x - y)\, d\tau_n(y).$$

Since σ'_1 is continuous and bounded, and since τ_n is a distribution function, it follows that ϕ_n has for every x a derivative ϕ'_n which may be obtained by differentiation beneath the integral sign, so that

$$\phi'_n(x) = \int_{-\infty}^{+\infty} \sigma'_1(x - y)\, d\tau_n(y).$$

Now $|x^1 - x^2| \leqq \delta$ implies that $|\sigma'_1(x^1 - y) - \sigma'_1(x^2 - y)| \leqq \omega(\delta)$ for every y, since $|x^1 - x^2| = |(x^1 - y) - (x^2 - y)|$. Hence

$$|\phi'_n(x^1) - \phi'_n(x^2)| \leqq \int_{-\infty}^{+\infty} \omega(\delta)\, d\tau_n(y) = \omega(\delta) \quad \text{whenever} \quad |x^1 - x^2| \leqq \delta.$$

This proves (i), since $\omega(\delta) \to 0$ as $\delta \to 0$. Similarly, since $|\sigma'_1| < V(\sigma'_1)$,

$$|\phi'_n(x)| \leqq \int_{-\infty}^{+\infty} V(\sigma'_1)\, d\tau_n(y) = V(\sigma'_1),$$

so that (ii) is satisfied by $C = V(\sigma'_1)$. In order to prove (iii), notice first

that, since $V(\sigma'_1) < +\infty$, one can write the above representation of $\phi'_n(x)$ in the form

$$\phi'_n(x) = [\tau_n(y)\sigma'_1(x-y)]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} \tau_n(y)\, d\sigma'_1(x-y),$$

where, as pointed out above, $\sigma'_1(-\infty) = 0$ and $\sigma'_1(+\infty) = 0$, while $\tau_n(y)$, being a distribution function, is bounded. Hence

$$\phi'_n(x) = -\int_{-\infty}^{+\infty} \tau_n(y)\, d\sigma'_1(x-y), \text{ i.e., } \phi'_n(x) = \int_{-\infty}^{+\infty} \tau_n(x-y)\, d\sigma'_1(y),$$

and so, for arbitrary m and x_i,

$$\sum_{i=1}^{m} | \phi'_n(x_i) - \phi'_n(x_{i-1}) | \leqq \int_{-\infty}^{+\infty} \sum_{i=1}^{m} | \tau_n(x_i - y) - \tau_n(x_{i-1} - y) | \, | d\sigma'_1(y) |.$$

Now let $x_0 < \cdots < x_m$. Then $x_0 - y < \cdots < x_m - y$ for every y; hence

$$\sum_{i=1}^{m} | \tau_n(x_i - y) - \tau_n(x_{i-1} - y) | \leqq V(\tau_n) = 1.$$

Consequently,

$$\sum_{i=1}^{m} | \phi'_n(x_i) - \phi'_n(x_{i-1}) | \leqq \int_{-\infty}^{+\infty} | d\sigma'_1(y) | = V(\sigma'_1) \text{ whenever } x_0 < \cdots < x_m,$$

and so $V(\phi'_n) \leqq V(\sigma'_1)$; i.e., (iii) is satisfied by $c = V(\sigma'_1)$.

It is clear from the proof that the theorem can be extended to the case of multi-dimensional distributions and also to the case where one requires for $\sigma = \sigma_1 * \sigma_2 * \cdots$ derivatives higher than the first.

THE JOHNS HOPKINS UNIVERSITY.

A REPRESENTATION OF STIELTJES INTEGRALS BY CONDITIONALLY CONVERGENT SERIES.*

By Fritz John.

In an earlier paper the author expressed $\int_0^1 f(x)\,dx$, where $f(x)$ is a function of bounded variation by a series of the form $\sum_{\nu=1}^{\infty} a_\nu f(\lambda_\nu)$, a_ν and λ_ν being certain constants not depending on f.[1] The a_ν and λ_ν contained an arbitrary rational parameter γ. These expressions for the integral of a function were generalized by H. Rademacher,[2] who assigned to γ arbitrary algebraic values, and also proved the validity of the expansion for all Riemann integrable functions.

In the present paper I am giving a similar representation for Riemann-Stieltjes integrals:

$$\int_1^2 f(x)\,d\psi(x) = \sum_{\nu=0}^{\infty} a_\nu f(c_\nu) \tag{1}$$

where the a_ν and c_ν are independent of f. The sequences a_ν and c_ν are not uniquely determined by $\psi(x)$. One might expect that the c_ν can be prescribed arbitrarily to a certain extent (e. g. so as to form an everywhere dense set in the interval $1 \leqq x \leqq 2$) and that then coefficients a_ν depending on ψ can be determined such, that (1) holds for all functions f of a certain class. In this paper a special expansion of this sort is given for the case, that f is of bounded variation and that ψ is continuous; the arguments c_ν have the fixed values $\nu/2^{[\log_2 \nu]}$, independent of ψ. The series is in general only conditionally convergent. By introducing $y = \psi(x)$ as variable of integration, one can obtain new expansions $\sum_{\nu=0}^{\infty} a_\nu f(\lambda_\nu)$ for the Riemann integral $\int_1^2 f(y)\,dy$, the coefficients a_ν and the arguments λ_ν depending on an arbitrary continuous, monotonic function $\psi(x)$.

It would be desirable, to prove the more general theorem, that every continuous linear operator for functions of bounded variation f (the total

* Received December 2, 1936.

[1] "Identitäten zwischen dem Integral einer Funktion und unendlichen Reihen," *Mathematische Annalen*, vol. 110, pp. 718-721.

[2] "Some remarks on F. John's identity," *American Journal of Mathematics*, vol. 58 (1936), pp. 169-176.

variation of f taken as its norm) can be represented by a series of the form $\sum_\nu a_\nu f(c_\nu)$.

In what follows we shall denote with $\bar{x}$ the function defined by

$$\bar{x} = \begin{cases} \dfrac{x}{2^{[\log_2 x]}} & \text{for } x > 0 \\ 1 & \text{for } x = 0. \end{cases}$$

THEOREM. *Let $f(x)$ be of bounded variation and $\psi(x)$ be continuous in* $1 \leqq x \leqq 2$. *Then*

$$\int_1^2 f(x)\,d\psi(x) = \sum_{\nu=0}^{\infty} a_\nu f(\bar{\nu}),$$

the coefficients a_ν being defined by the following conditions:

$$(2) \qquad a_{2\nu} = \tfrac{1}{2}a_\nu + \tfrac{1}{2}\big(\psi(\overline{2\nu-2}) - \psi(\overline{2\nu-1})\big) \text{ for positive integers } \nu,$$

$$(3) \qquad a_\nu = \tfrac{1}{2}\big(\psi(\bar{\nu}) - \psi(\overline{\nu-1})\big) \text{ for *odd* positive integers } \nu,$$

$$(4) \qquad a_0 = \psi(2) - \psi(1).$$

Proof. We have for $\nu = 0, 1, 2, \cdots$

$$(5) \qquad \overline{2\nu} = \bar{\nu}, \qquad 1 \leqq \bar{\nu} < 2.$$

Moreover, as $\psi(x)$ is continuous, there exists for every positive ϵ a $\delta(\epsilon)$ such that $|\psi(x_1) - \psi(x_2)| < \epsilon$ for $1 \leqq x_1 \leqq 2$, $1 \leqq x_2 \leqq 2$ and $|x_1 - x_2| \leqq \delta(\epsilon)$.

Let ν be an odd number $> 2/\delta(\epsilon) + 1$. Then $[\log_2 \nu] = [\log_2 (\nu - 1)]$ and

$$|\bar{\nu} - \overline{\nu - 1}| = \frac{1}{2^{[\log_2 \nu]}} \leqq \delta;$$

therefore $|a_\nu| \leqq \epsilon/2$.

Let ν be even and $> 2/\delta(\epsilon) + 1$. Then

$$[\log_2(\nu - 2)] = [\log_2(\nu - 1)],$$

$$|\overline{\nu - 2} - \overline{\nu - 1}| = \frac{1}{2^{[\log_2(\nu-1)]}} < \delta,$$

$$|\psi(\overline{\nu-2}) - \psi(\overline{\nu-1})| < \epsilon,$$

$$|a_\nu| = |\tfrac{1}{2}a_{\nu/2} + \tfrac{1}{2}\big(\psi(\overline{\nu-2}) - \psi(\overline{\nu-1})\big)| \leqq \tfrac{1}{2}|a_{\nu/2}| + \epsilon/2.$$

Thus for all $\nu > 2/\delta(\epsilon) + 1$

$$|a_\nu| \leqq \tfrac{1}{2}|a_{[\nu/2]}| + \epsilon/2.$$

Consequently we have, if M_n denotes $\underset{2^n \leqq \nu < 2^{n+1}}{\text{Maximum}}\, a_\nu$

$$M_{n+1} \leqq \tfrac{1}{2}M_n + \epsilon/2$$

for sufficiently large n; therefore $\limsup_{n\to\infty} M_n \leqq \epsilon$, and as ϵ was an arbitrary positive number, $\lim_{n\to\infty} M_n = 0$. Hence

$$\lim_{\nu\to\infty} a_\nu = 0. \tag{6}$$

Let for $n = 0, 1, 2, \cdots$ and for $1 \leqq x < 2$

$$s_n(x) = \sum_{2^n \leqq \nu \leqq 2^n x} a_\nu.$$

Then for positive n

$$s_n(x) = \sum_{2^{n-1} \leqq \nu \leqq 2^{n-1}x} (a_{2\nu} + a_{2\nu-1}) - a_{2^n-1} + \theta_n a_{[2^n x]},$$

where

$$\theta_n = \begin{cases} 1 \text{ if } [2^n x] \text{ is odd} \\ 0 \text{ if } [2^n x] \text{ is even.} \end{cases}$$

Now according to (2), (3)

$$\begin{aligned} a_{2\nu} + a_{2\nu-1} = \tfrac{1}{2}a_\nu + \tfrac{1}{2}(\psi(\overline{2\nu-2}) - \psi(\overline{2\nu-1})) \\ + \tfrac{1}{2}\bigl(\psi(\overline{2\nu-1}) - \psi(\overline{2\nu-2})\bigr) = \tfrac{1}{2}a_\nu. \end{aligned}$$

Hence

$$s_n(x) = \tfrac{1}{2}s_{n-1}(x) - a_{2^n-1} + \theta_n a_{[2^n x]}. \tag{7}$$

As according to (6) $|a_{2^n-1}|$ and $|a_{[2^n x]}|$ are less than $\epsilon/4$ for sufficiently big n, we have for those n

$$|s_n(x)| \leqq \tfrac{1}{2}|s_{n-1}(x)| + \epsilon/2.$$

From this we may conclude, that

$$\lim_{n\to\infty} s_n(x) = 0 \text{ uniformly in } x \text{ for } 1 \leqq x < 2. \tag{8}$$

It follows moreover from (7), that

$$\sum_{n=1}^{N} s_n(x) = \tfrac{1}{2}\sum_{n=0}^{N-1} s_n(x) - \sum_{n=1}^{N} a_{2^n-1} + \sum_{n=1}^{N} \theta_n a_{[2^n x]}. \tag{9}$$

Now for $n \geqq 1$

$$\begin{aligned} a_{2^n-1} &= \tfrac{1}{2}\bigl(\psi(\overline{2^n-1}) - \psi(\overline{2^n-2})\bigr) = \tfrac{1}{2}\bigl(\psi(\overline{2^n-1}) - \psi(\overline{2^{n-1}-1})\bigr) \\ \sum_{n=1}^{N} a_{2^n-1} &= \tfrac{1}{2}\psi(\overline{2^N-1}) - \tfrac{1}{2}\psi(\overline{0}); \end{aligned}$$

besides for odd $[2^n x]$

$$\begin{aligned} \theta_n a_{[2^n x]} &= \tfrac{1}{2}\bigl(\psi(\overline{[2^n x]}) - \psi(\overline{[2^n x - 1]})\bigr) \\ &= \tfrac{1}{2}\bigl(\psi(\overline{[2^n x]}) - \psi(\overline{[2^{n-1} x]})\bigr) \end{aligned}$$

and for even $[2^n x] : [2^n x] = 2[2^{n-1}x]$

$$\theta_n a_{[2^n x]} = 0 = \tfrac{1}{2}\left(\psi(\overline{[2^n x]}) - \psi(\overline{[2^{n-1}x]})\right)$$

Thus

$$\sum_{n=1}^{N} \theta_n a_{[2^n x]} = \tfrac{1}{2}\left(\psi(\overline{[2^N x]}) - \psi(\overline{[x]})\right) = \tfrac{1}{2}\left(\psi(\overline{[2^N x]}\right.$$

Substituting these expressions in (9), we obtain

$$\tfrac{1}{2}\sum_{n=0}^{N} s_n(x) = -\tfrac{1}{2}s_N(x) + s_0(x) - \tfrac{1}{2}\psi(\overline{2^N - 1}) + \tfrac{1}{2}\psi(\overline{0}) - \tfrac{1}{2}\psi$$

Now

$$s_0(x) = a_1 = \tfrac{1}{2}\left(\psi(1) - \psi(\overline{0})\right),$$

$$\psi(\overline{2^N - 1}) = \psi\left(2 - \frac{1}{2^{N-1}}\right),$$

$$\lim_{N\to\infty} \psi(\overline{2^N - 1}) = \psi(2),$$

$$\psi(\overline{[2^N x]}) = \psi\left(\frac{[2^N x]}{2^N}\right)$$

$\lim_{N\to\infty} \psi(\overline{[2^N x]}) = \psi(x)$ uniformly in x for $1 \leqq x < 2$. H
follows that

(10) $$\sum_{n=0}^{\infty} s_n(x) = \psi(x) - \psi(2) \text{ uniformly in } x \text{ for } 1 \leqq$$

Let the step function $e(y, x)$ be defined by

$$e(y, x) = \begin{cases} 1 & \text{if } y \leqq x \\ 0 & \text{if } y > x. \end{cases}$$

Then for $1 \leqq x < 2$

$$\sum_{\nu=1}^{n} a_\nu e(\bar{\nu}, x) = \sum_{\substack{\nu=1 \\ 1 \leqq \bar{\nu} \leqq x}}^{n} a_\nu$$

$$= \sum_{\mu=0}^{[\log_2 n]-1} \sum_{2^\mu \leqq \nu \leqq 2^\mu x} a_\nu + \sum_{2^{[\log_2 n]} \leqq \nu \leqq \min(n,}$$

$$= \sum_{\mu=0}^{[\log_2 n]-1} s_\mu(x) + \begin{cases} s_{[\log_2 n]}(x) & \text{if } \bar{n} > \\ s_{[\log_2 n]}(n) & \text{if } \bar{n} \leqq \end{cases}$$

Because of (8) and (10) it follows that

(11) $$\sum_{\nu=1}^{\infty} a_\nu e(\bar{\nu}, x) = \psi(x) - \psi(2) \text{ uniformly in}$$

for $1 \leqq x < 2$. As $\lim_{x\to 2} e(\bar{\nu}, x) = 1$ and $\lim_{x\to 2} \psi(x) = \psi(2$

$1 \leqq x \leqq 2$; i.e.

$$\sum_{\nu=1}^{\infty} a_\nu = 0. \tag{12}$$

Let now $f(x)$ be of bounded variation in $1 \leqq x \leqq 2$.

We first assume, that f is also continuous. Then

$$f(y) = -\int_1^2 e(y, x)\, df(x) + f(2).$$

Consequently

$$\sum_{\nu=1}^{n} a_\nu f(\bar{\nu}) = -\int_1^2 \sum_{\nu=1}^{n} a_\nu e(\bar{\nu}, x)\, df(x) + f(2) \sum_{\nu=1}^{n} a_\nu.$$

As $\sum_{\nu=1}^{\infty} a_\nu e(\nu, x)$ converges uniformly in x and $f(x)$ is of bounded variation, it follows using (11) and (12)

$$\begin{aligned}
\sum_{\nu=1}^{\infty} a_\nu f(\bar{\nu}) &= -\int_1^2 \big(\psi(x) - \psi(2)\big)\, df(x) \\
&= -\int_1^2 \psi(x)\, df(x) + \psi(2)\,\big(f(2) - f(1)\big) \\
&= \int_1^2 f(x)\, d\psi(x) - f(2)\psi(2) + f(1)\psi(1) \\
&\qquad + \psi(2)f(2) - \psi(2)f(1) \\
&= \int_1^2 f(x)\, d\psi(x) + f(1)\,\big(\psi(1) - \psi(2)\big).
\end{aligned}$$

This proves our theorem (cf. the definition of a_0) for the case that f is continuous.

Let $f(x)$ be of bounded variation in $1 \leqq x \leqq 2$ and have discontinuities at the points ξ_μ $(\mu = 1, 2, \cdots)$. Let

$$f(\xi_\mu + 0) - f(\xi_\mu - 0) = -c_\mu, \qquad f(\xi_\mu) - f(\xi_\mu - 0) = d_\mu.$$

Let moreover $\chi(x, y)$ be defined by

$$\chi(x, y) = \begin{cases} 0 \text{ if } x \neq y \\ 1 \text{ if } x = y. \end{cases}$$

Then $f(x) = f_1(x) + f_2(x) + f_3(x)$, where $f_1(x)$ is continuous and of bounded variation,

$$f_2(x) = \sum_{\mu=1}^{\infty} c_\mu e(x, \xi_\mu), \qquad f_3(x) = \sum_{\mu=1}^{\infty} d_\mu \chi(\xi_\mu, x),$$

the series $\sum_{\mu=1}^{\infty} | c_\mu |$ and $\sum_{\mu=1}^{\infty} | d_\mu |$ being convergent. It is sufficient to prove our theorem for $f = f_1$, $f = f_2$, $f = f_2$ separately. For $f = f_1$ it follows from our

previous considerations. As (1) holds according to (11) for $f(x) = e(x, \xi_\mu)$ and $\sum_{\nu=1}^{\infty} a_\nu e(\bar{\nu}, \xi_\mu)$ converges uniformly in μ and $\sum_\mu c_\mu$ converges absolutely, it follows, that (1) holds as well for $f = f_2(x)$. In order to prove (1) for $f = f_3(x)$, it is only necessary to prove, that

$$\sum_{\nu=0}^{\infty} a_\nu \chi(\nu, \xi_\mu) = 0 \tag{13}$$

uniformly in μ. But, as

$$\chi(\bar{\nu}, x) = \lim_{h \to +0} \big(e(\nu, x) - e(\nu, x - h) \big)$$

it follows from (11), that

$$\sum_{\nu=1}^{\infty} a_\nu \chi(\bar{\nu}, x) = \psi(x) - \psi(x - 0) = 0$$

uniformly in x for $1 \leqq x \leqq 2$.

Thus (1) is proved generally for any $f(x)$ of bounded variation.

Remarks. If there is an ϵ such that $|f(x)| > \epsilon$ for all x in $1 \leqq x \leqq 2$, then the series $\sum_{\nu=0}^{\infty} a_\nu f(\bar{\nu})$ is certainly not absolutely convergent, unless $\psi(x)$ is a constant.

For if that series would be absolutely convergent, then $\sum_\nu |a_\nu|$ would be convergent; consequently according to (11), (13)

$$\begin{aligned} \psi(x) - \psi(2) &= \sum_{\nu=1}^{\infty} a_\nu e(\bar{\nu}, x) = \sum_{\nu \text{ odd}}^{\nu} \sum_{s=0}^{\infty} a_{2^s\nu} e(\overline{2^s\nu}, x) \\ &= \sum_{\nu \text{ odd}}^{\nu} \sum_{s=0}^{\infty} a_{2^s\nu} e(\bar{\nu}, x) \\ &= \sum_{\nu \text{ odd}}^{\nu} e(\bar{\nu}, x) \sum_{s=0}^{\infty} a_{2^s\nu} \\ &= \sum_{\nu \text{ odd}}^{\nu} e(\bar{\nu}, x) \sum_{\mu=0}^{\infty} a_\mu \chi(\bar{\mu}, \bar{\nu}) = 0. \end{aligned}$$

II. Let $\psi(x)$ be monotonically increasing and continuous and let $\psi(2) = 2$, $\psi(1) = 1$. Then we obtain by substituting $t = \psi(x)$ in (1) the identity

$$\int_1^2 f(t)\, dt = \sum_{\nu=0}^{\infty} a_\nu f(\lambda_\nu)$$

valid for every function f of bounded variation, a_ν being defined by (2), (3), (4) and λ_ν denoting $\psi(\bar{\nu})$.

UNIVERSITY OF KENTUCKY.

NOTE ON THE DEFINITION OF FIELDS BY INDEPENDENT POSTULATES IN TERMS OF THE INVERSE OPERATIONS.*

By David G. Rabinow.

1. Introduction. The concept of a field involves two operations which are usually called addition and multiplication. When the definition of the field is given in terms of these operations, we say it is defined for the direct operations.[1] The inverse operations of these direct operations may be called subtraction and division. It is possible to define a field in terms of these inverse operations.[2] The inverse operation of multiplication, division, when used as the fundamental operation, allows us to define multiplication in two essentially distinct ways. The first involves the fact that $1/1/a = a$ under certain restrictions on a. We can then define $ab = a(1/1/b)$. This is the method used in the paper referred to in footnote 2. The second method is probably the more fundamental in that it is analogous to the definition of division in terms of multiplication. We shall develop the definition of a field in terms of addition and division where our treatment of division shall be this second method. Instead of using subtraction we shall use addition, since this will simplify the proofs somewhat and since the use of subtraction has been completely discussed in the paper referred to in footnote 2.

2. Postulates for a field and theorems deducible from them. Let us consider the following set of postulates in connection with the base $(K, +, \circ)$ where K is a class of elements $a, b, c, \cdots$ and $+, \circ$ are binary operations.

Postulate 1. a in K and b in K imply $a + b$ in K.

Postulate 2. If $a, b, c, a + b, b + c, (a + b) + c, a + (b + c)$ are in K, then $(a + b) + c = a + (b + c)$.

Postulate 3. There exists in K at least one element Z such that $a + Z = a$ for all a in K.

* Received December 28, 1936.

[1] E. V. Huntington, "Note on the definition of abstract groups and fields by sets of independent postulates," *Transactions of the American Mathematical Society*, vol. 6 (1905), pp. 181-193.

[2] D. G. Rabinow, "Independent sets of postulates for abelian groups and fields in terms of the inverse operations," *American Journal of Mathematics*, vol. 59 (1937), pp. 211-224.

Postulate 4. For each element a in K there exists at least one element a' in K such that $a + a' = Z$.

Postulate 5. a in K and b in K and $b \neq Z$ imply aob in K.

Postulate 6. If a, b, c, aob, aoc, $(aob)oc$, $(aoc)ob$ are in K, then $(aob)oc = (aoc)ob$.

Postulate 7. If a, b, c, $a + b$, aoc, boc, $(a + b)oc$ and $aoc + boc$ are in K, then $(a + b)oc = aoc + boc$.

Postulate 8. a in K and b in K and $a \neq Z$ imply the existence of an unique element x in K such that $xoa = b$.

Postulate 9. If a, b, aoa, bob are in K, then $aoa = bob$.

Postulate 10. There exist at least two distinct elements in K.

Note 1. If we wish we may remove the uniqueness requirement in Postulate 8 and insert as an additional postulate either Lemma 2 or Lemma 3 proven below. However for compactness and for simplification of independence proofs, as well as for reasons which will become apparent in Section 4, the present form of Postulate 8 is desirable.

Note 2. Throughout the subsequent work the terms Postulate, Theorem, Lemma and Definition will be referred to respectively by P, T, L, and D. Those theorems deducible from Postulates 1 through 4 shall be assumed as known and will be referred to as G.

LEMMA 1. *If $b \neq Z$, then $Zob = Z$.*

Let a be any element in K. Hence since $b \neq Z$, we have by P5, P3, and P7, $aob = (a + Z)ob = aob + Zob$. Whence by G, $Zob = Z$.

LEMMA 2. *If $a \neq Z$, $b \neq Z$, then $aob \neq Z$.*

Since $b \neq Z$, there exists by P8 an unique element x such that $xob = Z$. Hence by L1, x must be Z. If $aob = Z$, then a must be Z. But this is a contradiction. Hence $aob \neq Z$.

LEMMA 3. *If a, b, c, aoc, boc are in K and if $c \neq Z$ and if $aoc = boc$, then $a = b$.*

By P4 there exists an element a' such that $a + a' = Z$. By P5 $a'oc$ is in K. Hence by P1 $a'oc + boc = a'oc + aoc$. Whence by P7 $(a' + b)oc = (a + a')oc = Zoc$ by P4. Hence by L2, $a' + b = Z$. Therefore by G $a = b$.

At this stage we are in the position to define:

Definition 1. There exists an unique element $U = aoa \neq Z$. For, by P10 there exists in K at least one element $a \neq Z$. Hence by P5 aoa is in K and by L2 $aoa \neq Z$. By P9 this defines the unique element $U = aoa = bob$.

LEMMA 4. *If under the conditions of* P8 *the element b is also not equal to Z, then $x \circ b = a$.*

For, $b = x \circ a$. Hence by P5 $(x \circ a) \circ b = b \circ b = U$ by D1. Whence by P6 $(x \circ b) \circ a = U = a \circ a$ since $a \neq Z$. Therefore by L3 $x \circ b = a$.

We can now define the product (written ab) of any two elements a and b as follows:

Definition 2. 1. $b = Z$, then $ab = Z$.

Definition 2. 2. $b \neq Z$, then ab shall be the element x of P8, that is the element x satisfying the equation $x \circ b = a$. From these definitions we have immediately

THEOREM 1. *a in K and b in K imply ab in K.*

THEOREM 2. *If a, b, ab, ba are in K, then $ab = ba$.*

Case I. $b = Z$, then by D2. 1 $ab = Z$. If $a = Z$, then by D2. 1 $ba = Z$. If $a \neq Z$, then by D2. 2 and L2 $ba = Z$.

Case II. $a \neq Z$, $b \neq Z$. By D2. 2 $ab = x$ where $x \circ b = a$. By D2. 2 $ba = y$ where $y \circ a = b$ and by L4 $y \circ b = a$. Hence by L3 $x = y$.

THEOREM 3. *If $a, b, c, ab, bc, (ab)c, a(bc)$ are in K, then $(ab)c = a(bc)$.*

Case I. If $a = Z$ or $b = Z$ or $c = Z$, the theorem follows as in Case I of T2.

Case II. $a \neq Z, b \neq Z, c \neq Z$. By D2. 2 let $(ab)c = w$ where $w \circ c = ab$. Hence by D2. 2 $(w \circ c) \circ b = a$. Similarly let $p = a(bc) = (bc)a$ by T2 where $p \circ a = bc$ and hence $(p \circ a) \circ c = b$ by D2. 2. By P6 $b = (p \circ a) \circ c = (p \circ c) \circ a$. Hence by L4 $(p \circ c) \circ b = a$. Therefore $(w \circ c) \circ b = (p \circ c) \circ b$ and by repeated application of L3 $p = w$.

THEOREM 4. *If a, b, c, $a + b$, ac, bc, $(a + b)c$ and $ac + bc$ are in K, then $(a + b)c = ac + bc$.*

Case I. $c = Z$. By D2. 1 $(a + b)c = ac = bc = Z$; whence by P3 $(a + b)c = ac + bc$.

Case II. $c \neq Z$. By D2. 2 let $(a + b)c = p$ where $p \circ c = a + b$. Also let $ac = x$ where $x \circ c = a$ and let $bc = y$ where $y \circ c = b$. Then by P1 $a + b = x \circ c + y \circ c = (x + y) \circ c$ by P7. Hence by L3 $p = x + y$.

THEOREM 5. *If a, b, c, $a + b$, ca, cb, $c(a + b)$ and $ca + cb$ are in K, then $c(a + b) = ca + cb$.*

This theorem follows immediately from T4 and T2.

THEOREM 6. *There exists an unique element $u \neq Z$ such that $au = ua = a$ for all a in K.*

Consider the element $U \neq Z$ defined in D1.

Case I. $a = Z$. By D2. 1 $Ua = Z$ and by T2 $Ua = aU$.

Case II. $a \neq Z$. Let $Ua = p$ where $poa = U = aoa$ by D2. 2 and D1. Hence by L3 $p = a$. By T2 $aU = Ua$. Now suppose there exists another U' such that $aU' = U'a = a$. From $U'a = a$ we have by D2. 2 $aoa = U'$. But $aoa = U$. Hence $U = U'$, and U is unique.

THEOREM 7. *For any elements a and b in K where $a \neq Z$, there exists an unique element x such that $xa = ax = b$.*

Let $x = boa$. Then by D2. 2 $p = xa = (boa)a$ where $poa = boa$. Hence by L3 $p = b$ and by T2 $ax = xa$. Now suppose there exists another element x' such that $ax' = x'a = b$. Then by D2. 2 $x' = boa$. Hence $x' = x$.

THEOREM 8. *If a, b, $a + b$, $b + a$ are in K, then $a + b = b + a$.* Let d be any element of $K \neq Z$ and let D be the element x of T7 such that $dD = U$. By T4 and T5

$$(a + b)(d + d) = a(d + d) + b(d + d) = ad + ad + bd + bd.$$

Likewise by T4 and T5

$$(a + b)(d + d) = (a + b)d + (a + b)d = ad + bd + ad + bd.$$

Therefore

$$ad + ad + bd + bd = ad + bd + ad + bd.$$

Hence by P1, P2, P3, and P4 $ad + bd = bd + ad$ or by T5 $(a + b)d = (b + a)d$. Multiplying by D and using T1, T3, T7 and T6 we have $a + b = b + a$.

But P1, P2, P3, P4, T1, T2, T3, T4, T5, T6, T7, T8 are the postulates for a field in terms of the direct operations of addition and multiplication (Huntington). Hence any system $(K, +, \mathrm{o})$, which satisfies Postulates 1 through 10, is a field with respect to the direct operations of addition and multiplication. Furthermore from Theorem 7 we see immediately that the operation o is the inverse operation of multiplication. To complete the proof that our set of postulates is both a necessary and sufficient set to define a field we must show that from P1, P2, P3, P4, T1, T2, T3, T4, T5, T6, T7, T8 we can deduce P5, P6, P7, P8, P9, P10. For this purpose we define the operation o as follows:

Definition 3. If a is in K and b is in K and if $a \neq Z$, then $x = b \circ a$ shall be the element x in T7 satisfying $xa = b$.

LEMMA 5. *If a, b, c, ac, bc are in K and if $c \neq Z$ and if $ac = bc$, then $a = b$.*

For, by T7 the element c' exists such that $cc' = U$. Then by T1 $(ac)c' = (bc)c'$ or by T4, T7, T6 $a = b$.

THEOREM 9. *a in K, b in K and $b \neq Z$ imply $a \circ b$ in K.* (P5)
Follows immediately from D3.

THEOREM 10. *There exist at least two distinct elements in K.* (P10)
These are the elements Z and u whose existence is postulated in P3 and T6.

THEOREM 11. *a, b, $a \circ a$, $b \circ b$ in K imply $a \circ a = b \circ b$.* (P9)
By T6 $ua = a$ for all a. If $a \neq Z$, then by D3 $u = a \circ a$ for all a for which $a \circ a$ is in K.

THEOREM 12. *a in K, b in K, and $a \neq Z$ imply the existence of an unique element x in K such that $x \circ a = b$.* (P8)

Take $x = ba$. Then by D3 $x \circ a = b$ if $a \neq Z$. Now suppose there exists another element x' such that $x' \circ a = b$. Then by D3, $x' = ba$. Hence $x' = x$.

THEOREM 13. *If a, b, c, $a + b$, $(a + b) \circ c$, $a \circ c$, $b \circ c$, $a \circ c + b \circ c$ are in K, then $(a + b) \circ c = a \circ c + b \circ c$.* (P7)

By P1 and T9 $(a + b) \circ c$, $a \circ c$, $b \circ c$ are in K if $c \neq Z$. By T7 there exists an element x such that $xc = a + b$. Also by T7 there exists elements y and w such that $yc = a$ and $wc = b$. Hence $yc + wc = a + b = xc = (y + w)c$ by P1 and T4. Therefore by L5 $x = y + w$ and the theorem follows by D3.

THEOREM 14. *If a, b, c, $a \circ b$, $a \circ c$, $(a \circ b) \circ c$, $(a \circ c) \circ b$ are in K, then $(a \circ b) \circ c = (a \circ c) \circ b$.* (P6)

Case I. $a = Z$. By D3 $(a \circ b) \circ c = (a \circ c) \circ b = Z$ by T7.

Case II. $a \neq Z$. By T9 $a \circ b$, $a \circ c$, $(a \circ b) \circ c$, $(a \circ c) \circ b$ are in K if $b \neq Z$ and $c \neq Z$. Let $(a \circ b) \circ c = x$ and $(a \circ c) \circ b = w$. By D3 $xc = a \circ b$. Take $a \circ b = y$ where by D3 $yb = a$. Likewise $wb = a \circ c$ and $a \circ c = p$ where $pc = a$. Since $xc = y$, then $xcb = yb = a$ and since $wb = p$, then $wbc = pc = a$. Hence $(xc)b = (wb)c = (wc)b$ by T1, T2, T3. Hence by repeated applications of L5 $x = w$.

From the above we conclude that the set of Postulates 1 through 10 is both necessary and sufficient to define a field.

3. Independence of the postulates. The postulates are examined for independence by exhibiting examples of systems $(K, +, \circ)$ which fail to satisfy the correspondingly numbered postulates but satisfy the remaining postulates.

Example 1) K is the class of two elements $0, 1$ with $a+b$ and $a \circ b$ satisfying the following multiplication tables. The elements u, r, and s are elements not in K.

$a+b$	0	1
0	0	u
1	1	0

$a \circ b$	0	1
0	r	0
1	s	1.

Example 2) K is the class of all rational numbers, positive, negative, and zero. $a+b=a+2b$. $a \circ b = a/b$.

Example 3) K is the class of all positive rational numbers. $a+b=a+b$. $a \circ b = a/b$.

Example 4) K is the class of all positive rational numbers including zero. $a+b=a+b$. $a \circ b = a/b$.

Example 5) K is the class of all integers, positive, negative, and zero. $a+b=a+b$. $a \circ b = a/b$.

Example 6) K is the class of hypercomplex numbers of the form $\pi 1 + \omega i + \rho j$ where π, ω, ρ are rational numbers, positive, negative, and zero. $a+b=a+b$.

$$a \circ b = \frac{1}{\pi_2{}^2 + \rho_2{}^2 + (\omega_2 - \rho_2)^2} \times (\pi_1 1 + \omega_1 i + \rho_1 j)(\pi_2 1 + \omega_2 i + \rho_2 j),$$

where the product of the coefficients shall be the ordinary product of rational numbers and the "units" shall follow the table.

	1	i	j
1	1	$-i$	$-j$
i	i	1	-1
j	j	-1	2.

Example 7) K is the class of all integers, positive, negative and zero. $a+b=a+b$. $a \circ b = a-b$.

Example 8) K is the class of all integers, positive, negative and zero. $a+b=a+b$. $a \circ b = 0$.

Example 9) K is the class of all rational numbers, positive, negative and zero. $a + b = a + b$. $aob = ab$.

Example 10) K is the class consisting of the element 0 only. $a + b = a + b$. aob is undefined.

4. The concept of operational invariance. Let us consider the Postulates 1 through 8 inclusive. It is to be observed that the commutative postulate, that is $aob = boa$, cannot be deduced from these postulates. An independence example for this postulate would be: K is the class of all rational numbers, positive, negative and zero. $a + b = a + b$ and $aob = a$. We note further that if the commutative postulate is added to the set of Postulates 1 through 8, we obtain the definition of a field in terms of the direct operations, provided that we define $aoZ = Z$. In other words we have found a set of postulates (P1 through P8) which defines a field in terms of either the direct or the inverse operations depending on what additional postulates we desire to add to it. This naturally suggests the following problem: Suppose we consider the set P1 through P8 as a distinct set of postulates involving the operations $+$ and o. Let us further consider the inverse operation of o, which may be defined by means of P8. Call this operation X. Replace o, wherever it occurs in the set P1 through P8, by X. We now have a new set P1′ through P8′. The problem is can we from P1 through P8 deduce P1′ through P8′? The purpose of this section of the paper is to prove that this is true. Any system $(K, +, \mathrm{o})$, which has this property of replacing o by X, is to be defined as an operationally invariant system with respect to the operation o. It is clear that a field does not have this property but that there exists a subset of the postulates of the field, namely P1 through P8, which does. (This concept of operational invariance may obviously be extended to any type of system in which an inverse may be defined.) By P8 there exists a unique element x such that $xob = a$ if $b \neq Z$. This enables us to make the following definition:

Definition 4. $x = aXb$ if x is the element satisfying P8 when $b \neq Z$, that is, if x is the element such that $xob = a$. This proves

THEOREM 15. *a in K, b in K and $b \neq Z$ imply aXb in K.*

THEOREM 16. *If a, b, c, $a + b$, $(a + b)Xc$, aXc, bXc and $aXc + bXc$ are in K, then $(a + b)Xc = aXc + bXc$.*

By T15 aXc, bXc and $(a + b)Xc$ are in K if $c \neq Z$. Let $(a + b)Xc = w$ where by D4 $woc = a + b$. Also let $aXc = p$ and $bXc = q$ where again by D4 $poc = a$ and $qoc = b$. Hence by P1 $poc + qoc = a + b$ or by P7

$(p+q) o c = w o c$. Hence $p + q = w$ by L3. (It is to be noted that L1, L2 and L3 are still true since their proofs depended only on P1 through P8.)

THEOREM 17. *a in K and b in K and $a \neq Z$ imply the existence of an unique element w such that $wXa = b$.*

Take $w = boa$. Since $a \neq Z$, then by T15 $(boa)Xa = r$ where r is in K. Hence by D4 $roa = boa$. Whence by L3 $r = b$. Furthermore the element w must be unique since by D4 $w = boa$ which by P5 is uniquely determined by a and b.

LEMMA 6. *If a, b, aXb and $(aXb)ob$ are in K, then $(aXb)ob = a$.*

Let $w = aXb$. By D4 this means $wob = a$.

THEOREM 18. *If a, b, c, aXb, aXc, $(aXb)Xc$ and $(aXc)Xb$ are in K, then $(aXb)Xc = (aXc)Xb$.*

To satisfy the hypothesis of the theorem we see from D4 that $b \neq Z$ and $c \neq Z$. Take $(aXb)Xc = w$ where $woc = aXb$ by D4. Take also $(aXc)Xb = p$ where $pob = aXc$ by D4. Hence by P5 $(woc)ob = (aXb)ob$ and $(pob)oc = (aXc)oc$. By L6 $(aXb)ob = a$ and $(aXc)oc = a$. Hence $(woc)ob = (poc)ob$. Whence by L3 $p = w$.

5. Consistency of the postulates. To show that the set of postulates for a field is consistent we exhibit the set of all rational numbers with ordinary addition and division as our operations. To show the consistency of the set of postulates considered in section 4 we may take the set of all rational numbers with ordinary addition and either ordinary multiplication or ordinary division.

HARVARD UNIVERSITY.

A CORRECTION.

In my paper referred to in footnote 2, one of the postulates is incorrectly stated and several of the independence examples need to be restated.

On page 215, Postulate 18.2 should be: If a^* exists, then $a^* \neq z$ (provided $a \neq z$, $U \neq z$).

On page 223, Example 12 should be: K is the class of all positive rational numbers excluding zero. $a - b = a + b$. $aob = a/b$.

On page 223, Example 13 should have $a - b = | a - b |$.

On page 223, Example 17 should be: K is the class of all rational numbers, positive, negative, and zero. $a - b = a - b$. $aob = ab$.

On page 223, Example 18.2 should have $aob = a/b$ except $1/a = 0$, but $1/1 = 1$.

THE REPRESENTATION OF INTEGERS AS SUMS OF VALUES OF CUBIC POLYNOMIALS. II.*

By R. D. JAMES.

1. Introduction. In a previous paper under the same title[1] the author proved the following result.

THEOREM 1. *Let s be an integer $\geqq 9$ and let $P(x)$ be a polynomial of the form*

$$(1.1) \qquad P(x) = a(x^3 - x)/6 + b(x^2 - x)/2 + cx,$$

where a, b, c, are integers without a common factor, and $a \not\equiv 4c \pmod 8$. Then every sufficiently large integer is a sum of nine values of $P(x)$.

The condition $a \not\equiv 4c \pmod 8$ was an artificial one which could not be removed at the time. In the present paper we shall show that Theorem 1 is true without the restriction $a \not\equiv 4c \pmod 8$. The method of proof was suggested by two recent papers by L. K. Hua.[2] The new idea introduced by him may be explained briefly in the following way. If $\Phi(x)$ is a polynomial of degree k with integral coefficients, an integer θ was defined in § 2, I to be the highest power of a prime p which divided every coefficient of $\Phi'(x)$. Hua defines θ to be the highest power of a prime p for which $p^\theta | \Phi'(x)$ for all integers x. It may be shown that the two definitions are equivalent in the case of cubic polynomials except when $p = 2$. It is this difference when $p = 2$ which enables us to avoid the restriction $a \not\equiv 4c \pmod 8$ in Theorem 1.

The results obtained by Hua in the second of the papers to which reference was made above are correct, but there is an error at the beginning of § 16, page 45. The proofs which he gives are therefore not complete. He makes the following statement: — "Let θ be the highest power of a prime p such that $\Phi'(h) \equiv 0 \pmod{p^\theta}$ for all integers x." The $\Phi(h)$ which he is using is an integral-valued polynomial and not necessarily one with integral coefficients. Hence $\Phi'(h)$ need not be an integer and the congruence $\Phi'(h) \equiv 0$

* Received February 18, 1937.

[1] *American Journal of Mathematics*, vol. 56 (1934), pp. 303-315. This paper will be referred to as I.

[2] *American Journal of Mathematics*, vol. 58 (1936), pp. 553-562; *Journal of the Chinese Mathematical Society*, vol. 1 (1936), pp. 23-61.

(mod p^θ) is meaningless. In Lemmas 1 and 2 we shall show how to avoid this difficulty.

2. The proof of Hua's results. We first introduce the notation to be used. Let

$$P(x) = \sum_{j=0}^{k} a_j \frac{x(x+1)\cdots(x+j-1)}{j!},$$

where the a_j are integers. Let d be the least common multiple of the denominators of

$$a_0,\ a_1,\ a_2/2!,\ a_3/3!, \cdots, a_k/k!.$$

If the canonical product of an integer n is $n = p_1^{l_1}\cdots p_r^{l_r}$, and if $p_i^{a_i} \mid d$, $p_i^{a_i+1} \nmid d$ for $i = 1, 2, \cdots, r$, we define n^* by the equation $n^* = p_1^{l_1+a_1}\cdots p_r^{l_r+a_r}$.

Let $\Phi(x) = d\,P(x)$ so that $\Phi(x)$ is a polynomial with integral coefficients. For every prime p let θ be the highest power of p for which $\Phi'(x) \equiv 0 \pmod{p^\theta}$ for every integer x. Let $P_0(x) = p^{-\theta}\Phi'(x)$. Let $M(m)$ denote the number of solutions of

$$\sum_{\nu=1}^{s} P(x_\nu) \equiv n \pmod{m}, \qquad 0 \leqq x_\nu < m^*. \tag{2.1}$$

For $m = p^l$ let $N(p^l)$ denote the number of solutions of (2.1) in which not every $P_0(x_\nu)$ is divisible by p.

LEMMA 1. (Hua, Lemma 35). *If* $d = p^tD$, *where* $(p, D) = 1$, $l \geqq \max(2\theta + 2 - t, \theta + 2)$ *and* $x = y + dzp^{l-\theta-1}$, *then*

$$P(x) \equiv P(y) + zp^{l-1}P_0(y) \pmod{p^l},$$
$$P_0(x) \equiv P_0(y) \pmod{p}.$$

Proof. If we expand $\Phi(x) = \Phi(y + dzp^{l-\theta-1})$ by Taylor's Theorem we obtain

$$\begin{aligned}\Phi(x) &= \Phi(y) + zdp^{l-\theta-1}\Phi'(y) + \sum_{j=2}^{k} (zdp^{l-\theta-1})^j \frac{\Phi^{(j)}(y)}{j!}\\ &= \Phi(y) + zDp^{t+l-\theta-1}\Phi'(y) + \sum_{j=2}^{k} (zDp^{t+l-\theta-1})^j \frac{\Phi^{(j)}(y)}{j!}.\end{aligned}$$

Since $(t + l - \theta - 1)j \geqq 2(t + l - \theta - 1) \geqq t + l$ we have

$$\begin{aligned}\Phi(x) &\equiv \Phi(y) + zDp^{t+l-\theta-1}\Phi'(y) \pmod{p^{t+l}},\\ dP(x) &\equiv dP(y) + zDp^{t+l-1}P_0(y) \pmod{p^{t+l}},\\ p^tP(x) &\equiv p^tP(y) + zp^{t+l-1}P_0(y) \pmod{p^{t+l}},\\ P(x) &\equiv P(y) + zp^{l-1}P_0(y) \pmod{p^l}.\end{aligned}$$

This proves the first result of the lemma and the second follows in a similar way.

LEMMA 2. (Hua, Lemma 36). *If* $l \geqq \max(2\theta+2-t,\ \theta+2)$ *then* $N(p^l) = p^{s-1}N(p^{l-1})$.

Proof. The argument is very similar to that used in the proof of Lemma 3, I. If $d = p^tD$ where $(p, D) = 1$, then $p^{l*} = p^{t+l}$ so that $N(p^l)$ is the number of solutions of

$$(2.21) \qquad \sum_{\nu=1}^{s} P(x_\nu) \equiv n \pmod{p^l},\ 0 \leqq x_\nu < p^{t+l},\ p \nmid \text{every } P_0(x_\nu).$$

Hence $N(p^l)$ is equal to D^{-s} times the number of solutions of

$$(2.22) \qquad \sum_{\nu=1}^{s} P(x_\nu) \equiv n \pmod{p^l},\ 0 \leqq x_\nu < Dp^{t+l},\ p \nmid \text{every } P_0(x_\nu).$$

For every x_ν in (2.22) let $z_\nu = [x_\nu/(Dp^{t+l-\theta-1})]$ so that

$$x_\nu = y_\nu + z_\nu Dp^{t+l-\theta-1}, \qquad \begin{array}{l} 0 \leqq y_\nu < Dp^{t+l-\theta-1}, \\ 0 \leqq z_\nu < p^{\theta+1}. \end{array}$$

Then by Lemma 1 we may write (2.22) in the form

$$\begin{array}{l} \sum_{\nu=1}^{s} P(y_\nu) + p^{l-1} \sum_{\nu=1}^{s} z_\nu P_0(y_\nu) \equiv n \pmod{p^l}, \\ 0 \leqq y_\nu < Dp^{t+l-\theta-1},\ 0 \leqq z_\nu < p^{\theta+1},\ p \nmid \text{every } P_0(y_\nu). \end{array}$$

It follows that to each solution of (2.22) there corresponds a solution of the two congruences

$$(2.23) \qquad \sum_{\nu=1}^{s} P(y_\nu) \equiv n \pmod{p^{l-1}},\ \ 0 \leqq y_\nu < Dp^{t+l-\theta-1},\ \ p \nmid \text{every } P_0(y_\nu),$$

$$(2.24) \qquad \sum_{\nu=1}^{s} z_\nu P_0(y_\nu) \equiv p^{-l+1}\Big(n - \sum_{\nu=1}^{s} P(y_\nu)\Big) \pmod{p},\ \ 0 \leqq z_\nu < p^{\theta+1}.$$

By the same method of proof as that used in Lemma 3, I, it can be shown that (2.23) has $D^s p^{-\theta s} N(p^{l-1})$ solutions and that (2.24) has $p^{\theta s+s-1}$ solutions. Hence (2.22) has $p^{\theta s+s-1}D^s p^{-\theta s}N(p^{l-1})$ solutions. Then

$$N(p^l) = D^{-s}p^{\theta s+s-1}D^s p^{\theta s}N(p^{l-1}) = p^{s-1}N(p^{l-1}),$$

and this completes the proof.

3. The proof of Theorem 1 when $a \equiv 4c \pmod 8$. As explained in I, Theorem 1 is a consequence of Theorem 3, I, and this theorem in turn depends upon the fact that

$$p^{-l(s-1)}M(p^l) \geqq p^{-\gamma(s-1)}, \qquad l \geqq \gamma, \qquad s \geqq 9, \tag{3.11}$$

where γ is some fixed integer. If the proof of (3.11) as presented in I is examined, it is found that the restriction $a \not\equiv 4c \pmod 8$ was used only when $b \equiv 6c \pmod 8$, c odd. This was in (6.41). Hence in this section we shall prove that (3.11) is true when $P(x)$ has the form (1.1) with $a \equiv 4c$, $b \equiv 6c \pmod 8$ and c odd.

In sections 4 and 5 of I we have shown that the number of solutions of [3]

$$\sum_{\nu=1}^{s} P(vx_\nu + t) \equiv n \pmod{p^l}, \qquad 0 \leqq x_\nu < p^l \tag{3.12}$$

is $\geqq p^{(l-2)(s-1)}$ when $p \geqq 3$. Since $p^{l*} = p^l$ when $P(x)$ is a cubic polynomial and $p > 3$, it is evident that the number of solutions of (3.12) is the same as the number of solutions of

$$\sum_{\nu=1}^{s} P(y_\nu) \equiv n \pmod{p^l}, \qquad 0 \leqq y_\nu < p^{l*}. \tag{3.13}$$

The congruence (3.13) has $M(p^l)$ solutions by definition. Hence we have $M(p^l) \geqq p^{(l-2)(s-1)}$ when $p > 3$, and this proves (3.11) when $p > 3$. In a similar manner it can be shown that (3.11) is true with $\gamma = 2$ when $p = 3$ and $3 | a$, for in this case $v = 1$.

Thus two cases, $p = 3$, $3 \nmid a$ and $p = 2$, remain. We shall dispose of them in Lemmas 3 and 4.

LEMMA 3. *If $p = 3$, $3 \nmid a$ then (3.11) is true with $\gamma = 1$.*

Proof. In this case we have $d = 3$, $3^* = 9$, $\Phi(x) = 3P(x)$,

$$\Phi'(x) = 3a(x^2 + x)/2 + (6b - 3a)x/2 - (a + 3b - 6c)/2.$$

From this equation it is evident that $3 \nmid \Phi'(x)$ for all values of x since $3 \nmid a$. Hence $\theta = 0$ and $P_0(x) = \Phi'(x)$.

We distinguish two cases. 1). Suppose $3 | b$. Since

$$P(x+1) + 2P(9-x) + P(x-1) \equiv ax \pmod 3$$

there exists a solution x_0 of the congruence

$$P(x_0+1) + 2P(9-x_0) + P(x_0-1) \equiv n \pmod 3.$$

Hence the congruence

$$\sum_{\nu=1}^{5} P(x_\nu) \equiv n \pmod 3, \qquad 0 \leqq x_\nu < 3^*$$

[3] See (1.41)–(1.44) of I for the definition of v and t.

has the solution $x_1 = 0$, $x_2 = x_0 + 1$, $x_3 = x_4 = 9 - x_0$, $x_5 = x_0 - 1$. Also, $P_0(0) = -(a + 3b - 6c)/2$ and this expression is not divisible by 3. Hence for $s \geqq 5$, $l \geqq 2 = \max(2\theta + 2 - t, \theta + 2)$, it follows from Lemma 2 that

$$(3.2) \quad M(3^l) \geqq N(3^l) = 3^{s-1}N(3^{l-1}) = \cdots = 3^{(l-1)(s-1)}N(3) \geqq 3^{(l-1)(s-1)}.$$

This proves the lemma when $3 \mid b$.

2). Suppose $3 \nmid b$. In this case we have

$$P(x) + P(9 - x) \equiv bx^2 \pmod 3$$

and so there is a solution of the congruence [4]

$$\sum_{\nu=1}^{3} [P(x_\nu) + P(9 - x_\nu)] \equiv n \pmod 3.$$

As before we have a solution $y_1 = 0$, $y_{2\nu} = x_\nu$, $y_{2\nu+1} = 9 - x_\nu$, $\nu = 1, 2, 3$, of the congruence

$$\sum_{\mu=1}^{7} P(y_\mu) \equiv n \pmod 3, \qquad 0 \leqq y_\mu < 3^*,$$

in which $P_0(0)$ is not divisible by 3. Thus (3.2) is proved in this case also with $s \geqq 7$.

LEMMA 4. *If $p = 2$ then (3.11) is true with $\gamma = 5$.*

Proof. In this case we have $d = 1$ or 3,

$$\Phi(x) = da(x^3 - x)/6 + db(x^2 - x)/2 + dcx$$
$$\Phi'(x) = da(x^2 + x)/2 + d(2b - a)x/2 - d(a + 3b - 6c)/6.$$

Using the relations $a \equiv 4c$, $b \equiv 6c \pmod 8$ it is easily seen that $\Phi'(x) \equiv 0 \pmod 4$ for all values of x, but that $\Phi'(x) \not\equiv 0 \pmod 8$ for all values of x. Hence $\theta = 2$.

If n is even let $x_0 = 0$ or 1 according as $n \equiv 2$ or $n \equiv 0 \pmod 4$. Then since $P(0) = 0$ and $P(1) = c$ is odd, we have $n - 2P(x_0) \equiv 2 \pmod 4$. If n is odd let $r = 1$ or 3 according as $n \equiv 3c$ or $n \equiv c \pmod 4$. Then we have $n - rP(1) \equiv 2 \pmod 4$. This shows that we can always write n in the form

$$n = rP(x_0) + 2n_1,$$

where $r = 1, 2,$ or 3, and n_1 is odd.

[4] E. Landau, *Vorlesungen über Zahlentheorie,* Bd. I, Theorem 301.

Now $P(x) + P(32 - x) \equiv bx^2 \pmod{32}$ and since $b/2$ is odd there exists a solution of the congruence [5]

$$(b/2) \sum_{\nu=1}^{3} x_\nu^2 \equiv n_1 \pmod{16}$$

in which x_1 is odd. It follows that

$$\sum_{\nu=1}^{3} [P(x_\nu) + P(32 - x_\nu)] \equiv 2n_1 \pmod{32}, \qquad 0 \leqq x_\nu < 16$$

and hence that

$$rP(x_0) + \sum_{\nu=1}^{3} [P(x_\nu) + P(32 - x_\nu)] \equiv n \pmod{32}.$$

Moreover at least one of $P_0(x_1)$ and $P_0(32 - x_1)$ is not divisible by 2. For if both were divisible by 2 we should have

$$dbx_1/2 \equiv P_0(x_1) - P_0(32 - x_1) \equiv 0 \pmod{2},$$

whereas d, $b/2$, and x_1 are all odd. Then for

$$s \geqq 9 \geqq r + 6, \qquad l \geqq 6 = \max(2\theta + 2 - t, \theta + 2),$$

we have

$$M(2^l) \geqq N(2^l) = 2^{s-1}N(2^{l-1}) = \cdots = 2^{(l-5)(s-1)}N(32) \geqq 2^{(l-5)(s-1)}.$$

This completes the proof for the case $p = 2$.

THE UNIVERSITY OF CALIFORNIA,
BERKELEY, CALIFORNIA.

[5] E. Landau, *Vorlesungen über Zahlentheorie*, Bd. I, Theorem 301.

THE CONJUNCTIVE EQUIVALENCE OF PENCILS OF HERMITIAN AND ANTI-HERMITIAN MATRICES.*

By John Williamson.

Let K be a commutative field of characteristic zero and let $K(i)$ be a quadratic adjunction field of K, where i is a zero of the polynomial $x^2 - a$, irreducible in K. If A is a matrix with elements in $K(i)$, or more shortly a matrix over $K(i)$, the matrix A^* is defined to be the conjugate transposed of A, so that $A^* = \bar{A}'$. In particular, if A is a matrix over K, $A^* = A'$, the transposed of A. Let

$$(1) \qquad \Lambda = rA + sB,$$

be a pencil of matrices, in which

$$(2) \qquad A^* = \epsilon A, \qquad B^* = \delta B, \qquad \epsilon, \delta = \pm 1,$$

so that A is either hermitian or anti-hermitian and so is B. Let $\Lambda_1 = rA_1 + sB_1$ be another such pencil. Then the two pencils Λ and Λ_1 are said to be *conjunctively equivalent,* if there exists a non-singular matrix P over $K(i)$ such that

$$P\Lambda P^* = \Lambda_1;$$

that is, if $PAP^* = A_1$ and $PBP^* = B_1$. When the matrices A, B, A_1, B_1 are all matrices over K, the two pencils are said to be *congruently equivalent,* if there exists a non-singular matrix P over K, such that

$$P\Lambda P' = \Lambda_1.$$

There are accordingly two distinct problems to be considered; (a) to determine necessary and sufficient conditions that two pencils, which satisfy (2), be conjunctively equivalent and (b) to determine necessary and sufficient conditions that two such pencils be congruently equivalent. Problem (a) has been solved completely, for the case in which $\epsilon = \delta = 1$, when K is the field of all real numbers [1] and also, under the restriction that B be non-singular, when K is a commutative field of characteristic zero.[2] In the other two cases,

* Received November 30, 1936.

[1] H. W. Turnbull, "On the equivalence of pencils of hermitian forms," *Proceedings of the London Mathematical Society,* vol. 39 (1935), pp. 232-248; M. H. Ingraham and K. W. Wegner, "The equivalence of pairs of hermitian matrices," *Transactions of the American Mathematical Society,* vol. 38 (1935), pp. 145-162. In both papers a treatment of singular pencils is given.

[2] John Williamson, "The equivalence of non-singular pencils of matrices in an

in which one or both of ϵ, δ have the value -1, problem (a) may be reduced to the case in which $\epsilon = \delta = 1$. For example, let $\epsilon = -1$, $\delta = 1$, so that $A = -A^*$, $B = B^*$. Then, $(iA)^* = iA$ and

$$\Lambda = tC + sB, \tag{3}$$

where $t = r/i$, $C = C^*$, $B = B^*$ and (3) is a pencil of hermitian matrices. Hence, when Λ is a non-singular pencil, problem (a) is completely solved.

Problem (b) has been solved for the case, in which $\epsilon = \delta = 1$, when K is the real field [3] and also, under the restriction that B be non-singular, when K is a general commutative field of characteristic zero, (I). It is however not possible to reduce the other cases to this one and they must be considered separately. The problem has also been solved for a non-singular pencil and general field K, when $\epsilon = 1$, $\delta = -1$.[4] The remaining problem when $\epsilon = \delta = -1$, so that both A and B are skew symmetric matrices, is considered here. It is shown that two skew symmetric matrices are congruently equivalent, if they have the same kronecker minimal indices and the same invariant factors.[5] This is a much simpler result than those obtained in the other cases; but it is only natural that this be so, since two skew symmetric matrices of the same rank are congruently equivalent, while the same is not true of two symmetric matrices.

Since, when A and B are both skew symmetric matrices of odd order, every matrix of the pencil Λ is singular, a treatment of singular pencils is absolutely necessary. Accordingly the conjunctive or congruent equivalence of two general pencils Λ, which satisfy (2), is first considered, so that the solution of both problems (a) and (b) is completed in all cases. The method here adopted is quite distinct from that used in the discussion of non-singular pencils of hermitian matrices,[6] and has the advanatge that at no stage is a change made in the basis of the pencil.

Section 1 is devoted to the proofs of subsidiary lemmas, section 2 to the consideration of singular pencils, section 3 to that of non-singular pencils in which B is singular, and 4 to the reduction of a non-singular pencil of skew symmetric matrices to canonical form.

---

arbitrary field," *American Journal of Mathematics*, vol. 57 (1935), pp. 475-490. This paper will be referred to as I.

[3] Turnbull, *loc. cit.*

[4] John Williamson, "On the algebraic problem concerning the normal forms of linear dynamical systems," *American Journal of Mathematics*, vol. 58 (1936), pp. 141-163. This paper will be referred to as II.

[5] This is a well known result in case K is algebraically closed. See L. E. Dickson, *Modern Algebraic Theories*, p. 125, or C. C. MacDuffee, *The Theory of Matrices*, p. 61.

[6] Turnbull, *loc. cit.*; Wegner and Ingraham, *loc. cit.*

1. If $\Lambda = rA + sB$, we define the matrix pencil Λ'' by

$$\Lambda'' = \epsilon r A^* + \delta s B^*. \tag{4}$$

Let ρ_j be the matrix pencil

$$\rho_j = \begin{pmatrix} r & s & 0 & \cdot & 0 & 0 \\ 0 & r & s & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & s & 0 \\ 0 & 0 & 0 & \cdot & r & s \end{pmatrix}, \tag{5}$$

of j rows and $j+1$ columns. Then ρ_j'' is a matrix pencil of $j+1$ rows and j columns, while the matrix pencil

$$R_j = \begin{pmatrix} \rho_j & 0 \\ 0 & \rho_j'' \end{pmatrix} \tag{6}$$

is a square matrix of $2j+1$ rows and columns. Let

$$N_j = rE_j + sU_j, \tag{7}$$

where E_j and U_j are respectively the unit matrix and the auxiliary unit matrix of order j.[7] We now prove a few elementary lemmas involving the matrices R_j and N_j.

LEMMA I. *Let S and T be two matrices, which satisfy*

$$SR_q = R_p''T. \tag{8}$$

Then, if $p < q$, the first row of S is zero; if $p = q$,

$$S = \begin{pmatrix} 0 & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}, \tag{9}$$

where σ_{12} and σ_{21} are diagonal matrices of orders $p+1$ and p respectively.

Proof. Since S and T satisfy (8) we may write S and T as two rowed square matrices of matrices,

$$S = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}, \qquad T = \begin{pmatrix} \tau_{11} & \tau_{12} \\ \tau_{21} & \tau_{22} \end{pmatrix},$$

where σ_{11} is a matrix of $p+1$ rows and q columns, σ_{12} a matrix of $p+1$ rows and $q+1$ columns etc. From (8) we immediately deduce the four equations

$$\sigma_{11}\rho_q = \rho_p''\tau_{11}, \quad \sigma_{12}\rho_q'' = \rho_p''\tau_{12}, \quad \sigma_{21}\rho_q = \rho_p\tau_{21}, \quad \sigma_{22}\rho_q'' = \rho_p\tau_{22}. \tag{10}$$

[7] See Turnbull and Aitken, *Canonical Matrices*, p. 62.

The first of these equations is of the form,

$$C\rho_q = \rho_p{}''D, \tag{11}$$

where $C = (c_{ij})$ is a matrix of $p+1$ rows and q columns, while $D = (d_{ij})$ is a matrix of p rows and $q+1$ columns. As a consequence of (11) we have

$$rc_{ij} + sc_{i,j-1} = \epsilon r d_{ij} + \delta s d_{i-1,j}, \quad (i = 1, 2, \cdots, p+1,\ j = 1, 2, \cdots, q+1),$$

with the understanding that

$$c_{i,0} = d_{0,j} = c_{i,q+1} = d_{p+1,j} = 0. \tag{12}$$

Therefore,

$$c_{ij} = \epsilon d_{ij}, \quad c_{i,j-1} = \delta d_{i-1,j}, \quad c_{ij} = \epsilon\delta c_{i+1,j-1}, \tag{13}$$

and we see from the last of these equations that the elements in any counter diagonal of C are the same except perhaps for sign. But, as a consequence of (12) and (13),

$$c_{p+1,j} = \epsilon d_{p+1,j} = 0$$

and also

$$c_{i-1,1} = \epsilon\delta c_{i,0} = 0, \qquad (i = 2, 3, \cdots, p+1).$$

Hence the first column and the last row of C are zero, and consequently C is zero. Therefore σ_{11} is zero.

The second of equations (10) is of the form $C\rho_q{}'' = \rho_p{}''D$, where $C = (c_{ij})$ is a matrix of $p+1$ rows and $q+1$ columns, while $D = (d_{ij})$ is a matrix of p rows and q columns. Consequently

$$c_{ij} = \epsilon d_{ij}, \quad c_{i,j+1} = \delta d_{i-1,j}, \quad c_{ij} = \epsilon\delta c_{i+1,j+1},$$
$$(i = 1, 2, \cdots, p+1;\ j = 1, 2, \cdots, q),$$

so that the elements in any diagonal of C differ at most in sign. But

$$c_{1,j+1} = \delta d_{0,j} = 0 \quad \text{and} \quad c_{p+1,j} = \epsilon d_{p+1,j} = 0.$$

Hence

$$c_{1j} = 0, \qquad (j = 2, \cdots, q+1),$$
$$c_{p+1,j} = 0, \qquad (j = 1, 2, \cdots, q).$$

Therefore, if the number of rows of C is less than the number of columns, i. e. if $p < q$, $C = 0$ and, if $p = q$, C is a diagonal matrix. Hence, if $p < q$, $\sigma_{12} = 0$ and, if $p = q$, σ_{12} is a diagonal matrix. Moreover, if $p = q$, D is also a diagonal matrix, whose first element is ϵ times the first element of C. We have as a result the corollary:

COROLLARY I. *If $T = S^*$ and σ_{12} is non-singular, σ_{21} is non-singular.*

The proofs of the following lemmas, which are similar to the above, are omitted.

LEMMA 2. *If* $SN_i = R_q''T$, *the first row of* S *is zero.*

LEMMA 3. *If* M *is a square matrix of order* t *and* $S(rM + sE_t) = R_q''T$ *the first row of* S *is zero.*

LEMMA 4. *If* $SN_i = N_j''T$ *and* $i < j$, *the first row of* S *is zero.*

LEMMA 5. *If* $S(rM + sE) = N_i''T$, *the first row of* S *is zero.*

When G is a square matrix of order n, we may consider G as a matrix of matrices and write

$$G = (G_{ij}), \qquad (i, j = 1, 2, \cdots, t),$$

where G_{ij} is a matrix of n_i rows and n_j columns. If H is a second n-rowed square matrix

$$H = (H_{ij}), \qquad (i, j = 1, 2, \cdots, t),$$

where H_{ij} is also a matrix of n_i rows and n_j columns we shall say that G and H are *similarly partitioned.* If $G_{ij} = 0$, when $i \neq j$, we shall call G a *diagonal block* matrix and write

$$G = [G_{11}, G_{22}, \cdots, G_{tt}].$$

2. Let Λ be the pencil defined by (1) and (2) so that

$$\Lambda = \Lambda''. \tag{14}$$

Let Λ annihilate the column vector u of dimension n, whose elements are polynomials in r and s with coefficients in $K(i)$. Then

$$\Lambda u = 0, \text{ identically in } r \text{ and } s,$$

and consequently

$$0 = u''\Lambda'' = u''\Lambda, \text{ by } (14).$$

Since u'' is the row vector obtained from u^* by replacing r by ϵr and s by δs, the degree of u'' in r and s is the same as the degree of u. Therefore *the set of kronecker minimal row indices*[8] *coincides with the set of minimal column indices.* In particular, if $u_1, u_2, \cdots, u_p$ form a complete set of linearly independent vectors over $K(i)$ annihilated by Λ and, if V is any non-singular matrix, whose first p columns are the vectors $u_1, u_2, \cdots, u_p$, then

$$V^*\Lambda V = [0, \Lambda_1], \tag{15}$$

[8] Turnbull and Aitken, *op. cit.*, pp. 119-125.

where Λ_1 is a pencil of the same type as Λ but of order $n-p$. Moreover none of the minimal row (or column indices) of Λ_1 is zero. Since (15) is a conjunctive transformation and since, when A and B are matrices over K, $V^* = V'$, we may consider the pencil Λ_1 instead of Λ. Accordingly without any loss of generality we may assume that *no minimal row or column index of Λ is zero.*

Let Π be any pencil equivalent to Λ in the more general sense that there exist two non-singular matrices Q and P, such that

$$Q\Lambda P = \Pi.$$

Then,

$$P^*\Lambda P = P^*Q^{-1}\Pi = H\Pi,$$

so that the pencil $H\Pi$ is conjunctively equivalent to Λ. Accordingly, as a consequence of (14),

(16) $$H\Pi = \Pi'' H^*.$$

We now prove

LEMMA 6. *Let $\Pi = [\Pi_1, \Pi_2]$ be a diagonal block matrix, where Π_i is of order n_i, and let the matrix H_1, formed from the first n_1 rows and columns of H, be non-singular. Then, if equation* (16) *is satisfied, there exists a non-singular matrix W such that*

$$W^*H\Pi W = [H_1\Pi_1, H_2\Pi_2],$$

is also a diagonal block matrix.

Proof. Let $H = (H_{ij})$, $i, j = 1, 2$, be a partition of H similar to that of Π. Then $H_{11} = H_1$ is non-singular.

Equation (16) implies

$$H_{ij}\Pi_j = \Pi_i'' H^*_{ji}, \qquad (i, j = 1, 2),$$

and consequently

$$\Pi_1 (H_{11}^{-1})^* H^*_{21} = H_{11}^{-1}\Pi_1'' H^*_{21} = H_{11}^{-1} H_{12}\Pi_2.$$

It now follows by a simple calculation that, if

$$W = \begin{pmatrix} E_1 & 0 \\ -H_{21}H_{11}^{-1} & E_2 \end{pmatrix},$$

where E_i is the unit matrix of order n_i,

$$W\Pi W^* = [H_1, H_2][\Pi_1, \Pi_2].$$

Since the matrix W depends solely on H we have the

COROLLARY. *If H is a matrix over K, W is a matrix over K and $W^* = W'$.*

We now take Π in the canonical form [9]

$$\Pi = [R_{j_1}, R_{j_2}, \cdots, R_{j_k}, N_{t_1}, N_{t_2}, \cdots, N_{t_m}, rM + sE]. \tag{17}$$

In (17) each matrix R_{j_i} is defined by (6) and $j_1, j_2, \cdots, j_k$ form a complete set of minimal row (or column) indices, each matrix N_{t_i} is defined by (7) and corresponds to an elementary factor r^{t_i} of Λ; and $rM + sE$ is a pencil whose second member is non-singular and can therefore be taken in this simplified form.

For convenience we write (17) in the form

$$\Pi = [\pi_1, \pi_2, \cdots, \pi_{k+m+1}], \tag{18}$$

where $\pi_i = R_{j_i}$, $i < k$; $\pi_{k+i} = N_{t_i}$, $i < m$; $\pi_{k+m+1} = rM + sE$.

If $H = (H_{pq})$, $p, q = 1, 2, \cdots, k + m + 1$, is a partition of H similar to that of Π in (18), it follows from (16) that

$$H_{pq}\pi_q = \pi_p'' H^*_{qp}. \tag{19}$$

Let the integers $j_1, j_2, \cdots, j_k$ in (17) be so arranged that

$$j_1 = j_2 = \quad = j_c < j_{c+1} \leqq \cdots \leqq j_k.$$

If H_{11} is singular and, for some value of $p \leqq c$, H_{pp} is non-singular, by a conjunctive transformation involving an interchange of rows and the same interchange of columns, we may interchange H_{11} and H_{pp} without disturbing Π. By (19) and Lemmas (1), (2) and (3), the first row of H_{pq} is zero, when $p \leqq c$ and $q > c$, and contains at most one element h_{pq}, in the $(j_1 + 1)$-th place, different from zero when $p \leqq c$ and $q \leqq c$. Further, by the corollary to Lemma (1), h_{pp} is zero, if, and only if, H_{pp} is singular. Therefore, if H_{11} is singular, $h_{11} = 0$, and since H is non-singular at least one element in the first row of H is different from zero. Accordingly, for at least one value of $p \leqq c$, $h_{1p} \neq 0$. Without any loss of generality, then, we may assume that the first row of H_{12} contains the element h_{12} distinct from zero.

Let E be the unit matrix of order $2j + 1 = 2j_1 + 1$ and let

$$W_1 = \begin{pmatrix} E & E \\ 0 & E \end{pmatrix}, \qquad W_2 = \begin{pmatrix} E & iE \\ 0 & E \end{pmatrix}. \tag{20}$$

---

[9] W. Ledermann, "Reduction of singular pencils of matrices," *Proceedings of the Edinburgh Mathematical Society*, vol. 4 (1935), series 2, pp. 92-105.

Then

$$W_1(H_{pq})[R_j, R_j]W^*_1 = (G_{pq})[R_j, R_j],$$
$$W_2(H_{pq})[R_j, R_j]W^*_2 = (F_{pq})[R_j, R_j], \qquad (p, q = 1, 2),$$

where

$$G_{11} = H_{11} + H_{12} + H_{21} + H_{22}, \qquad F_{11} = H_{11} + iH_{21} - iH_{12} - i^2H_{22}.$$

Let g_{11} and f_{11} be the elements in the first row and $(j+1)$-th column of the matrices G_{11} and F_{11} respectively. Then

(21) $\quad g_{11} = h_{12} + h_{21} + h_{11} + h_{22}, \quad f_{11} = i(h_{21} - h_{12}) + h_{11} - i^2h_{22}.$

If H_{11} and H_{22} are both singular, $h_{11} = h_{22} = 0$ and at least one of g_{11} or f_{11} is non-zero, as otherwise h_{12} would be zero. Therefore at least one of G_{11} or F_{11} is non-singular.

If, however, we are restricted to congruent transformations over the field K, we are not at liberty to use the transformation W_2. But, if

$$Q = [-E_{j+1}, E_j, E_{2j+1}],$$
$$[R_j, R_j]Q = [-E_j, E_{j+1}, E_{2j+1}][R_j, R_j],$$

and

$$Q(H_{pq})[R_j, R_j]Q' = (K_{pq})[R_j, R_j], \qquad (p, q = 1, 2),$$

where

$$K_{12} = [-E_{j+1}, E_j]H_{12}, \qquad K_{21} = H_{21}[-E_j, E_{j+1}].$$

Accordingly, if $W_2 = W_1Q$, equation (21) becomes

$$g_{11} = h_{12} + h_{21}, \qquad f_{11} = h_{21} - h_{12},$$

so that once again either G_{11} or F_{11} is non-singular. Therefore, we may suppose that the necessary transformation has already been made and that H_{11} is non-singular. As a consequence of Lemma (6), there exists a non-singular matrix P_1, such that

$$P_1H\Pi P^*_1 = [H_{11}R_{j_1}, H_2\Pi_2].$$

By repeating this process k times we prove the existence of a non-singular matrix P such that

(22) $$PH\Pi P^* = [H_1R_{j_1}, H_2R_{j_2}, \cdots, H_kR_{j_k}, H_{k+1}\Pi_{k+1}],$$

where $\Pi_{k+1} = [N_{i_1}, N_{i_2}, \cdots, N_{i_t}, rM + sE]$. All matrices H_i in (22) are non-singular and satisfy the equations

(23) $\quad H_iR_{j_i} = R_{j_i}''H^*_i, \ (i = 1, 2, \cdots, k), \quad H_{k+1}\Pi_{k+1} = \Pi''_{k+1}H^*_{k+1}.$

To simplify equations (23) we write $H_i = S$, $R_{j_1} = R_p$, and with the notation of Lemma 1, have

$$S = \begin{pmatrix} 0 & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix},$$

where σ_{12} and σ_{21} are non-singular and satisfy

$$(24) \qquad \sigma_{12}\rho_p'' = \rho_p''\sigma^*_{21}, \qquad \sigma_{22}\rho_p'' = \rho_p\sigma^*_{22}.$$

Let W be the matrix

$$W = \begin{pmatrix} \sigma_{12}^{-1} & 0 \\ -\frac{1}{2}\sigma_{22}\sigma_{12}^{-1} & E_p \end{pmatrix}.$$

Then,

$$R_pW^* = \begin{pmatrix} \rho_p(\sigma_{12}^{-1})^* & -\frac{1}{2}\rho_p(\sigma_{12}^{-1})^*\sigma^*_{22} \\ 0 & \rho_p'' \end{pmatrix} = \begin{pmatrix} \sigma_{21}^{-1}\rho_p & -\frac{1}{2}\sigma_{21}^{-1}\sigma_{22}\rho_p'' \\ 0 & \rho_p'' \end{pmatrix} \text{ by } (24)$$

$$= \begin{pmatrix} \sigma_{21}^{-1} & -\frac{1}{2}\sigma_{21}^{-1}\sigma_{22} \\ 0 & E_{p+1} \end{pmatrix} R_p,$$

and

$$WSR_pW^* = \begin{pmatrix} 0 & E_{p+1} \\ E_p & 0 \end{pmatrix} R_p.$$

Therefore each matrix H_i in (22) may be reduced to the form I_{j_i}, where

$$(25) \qquad I_j = \begin{pmatrix} 0 & E_{j+1} \\ E_j & 0 \end{pmatrix}.$$

Accordingly we have proved,

THEOREM I. *There exists a non-singular matrix P such that*

$$(26) \qquad P\Lambda P^* = [I_{j_1}R_{j_1}, I_{j_2}R_{j_2}, \cdots, I_{j_k}R_{j_k}, H_{k+1}\Pi_{k+1}],$$

where I_j is defined by (25) *and R_j by* (6), *while the pencil $H_{k+1}\Pi_{k+1}$ is non-singular.*

COROLLARY 1. *If the elements of A and B lie in K, the matrix P lies in K and $P^* = P'$.*

COROLLARY 2. *The canonical form on the right of* (26) *is determined completely by the minimal indices of Λ and the non-singular core $H_{k+1}\Pi_{k+1}$.*

3. In considering the reduction of the non-singular case we could, by a change of the basis of the pencil, reduce it to one in which the second member is non-singular. As a change of basis is quite distinct in nature from the conjunctive transformations of the pencil it is more satisfactory to proceed as follows.[10] For simplicity we write

$$(27) \qquad \Pi_{k+1} = \Pi = [N_{t_1}, N_{t_2}, \cdots, N_{t_m}, rM + sE],$$

and

$$(28) \qquad H_{k+1} = H = (H_{ij}), \qquad (i, j = 1, 2, \cdots, m+1),$$

---

[10] Cf. Ledermann, *loc. cit.*

where (28) is a partition of H similar to that of Π in (27). If

$$(29) \qquad t_1 = t_2 = \quad = t_c > t_{c+1} \geqq \cdots \geqq t_m,$$

it follows from Lemmas (4) and (5) that the first row of $H_{1p} = 0$, when $p > c$, and that at most one element in the first row of H_{1q} is different from zero, when $p \leqq c$. If H_{11} is singular, we may suppose, as in the previous section, that H_{12} is non-singular and, since $H_{12} = \epsilon H^*_{21}$, that $\begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix}$ is non-singular. Therefore by several applications of Lemma 6 we may reduce the pencil $H\Pi$ to a diagonal block form where each block is either of type (α) $H_{11}N_t$ or of type (β) $\begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} \begin{pmatrix} N_t & 0 \\ 0 & N_t \end{pmatrix}$.

It is now necessary to consider three distinct cases;

(1) $\epsilon = \delta = 1$, (2) $\epsilon = +1$, $\delta = -1$, (3) $\epsilon = \delta = -1$.

Case 1. Both matrices in the pencil are hermitian or both symmetric. A block of type (β) may be reduced to two of type (α) (I page 481). The matrix H_{11} in (α) is of the form,

$$H_{11} = T_t(g_1 E_t + g_2 U_t + \cdots + g_{t-1} U_t^{t-1}),$$

where T_t is the counter unit matrix of order t.

Finally H_{11} may be reduced to the form

$$(30) \qquad H_{11} = g_1 T_t,$$

where g_1 lies in K.

In (30) g_1 is not uniquely determined. In fact, the diagonal matrix $G = [g_1, g_2, \cdots, g_c]$, where c is defined by (29), may be replaced by the diagonal matrix $F = [f_1, f_2, \cdots, f_c]$, provided F and G are conjunctively equivalent (I pages 482-487), or

THEOREM 2. *If r^t occur exactly c times among the elementary factors of a pencil of hermitian matrices Λ, in the canonical form for Λ there is a block $[g_1T_t, g_2T_t, \cdots, g_cT_t]$. The diagonal matrix $[g_1, g_2, \cdots, g_c]$ is determined apart from a conjunctive transformation.*

Theorems 1 and 2 together with the theorem in I, page 487, give a complete solution of problem (a).

Case 2. As stated in the introduction, it is only necessary to consider congruent transformations of pencils with elements in K. The matrix H_{11} is symmetric, while the matrix $H_{11}U_t$ is skew symmetric. Therefore

$$(31) \qquad H_{11}U = -U'H'_{11} = -U'H_{11}, \text{ where } U = U_t.$$

Let

$$X = \begin{pmatrix} 0 & \cdot & \cdot & \cdot & 0 & 1 \\ 0 & \cdot & \cdot & \cdot & -1 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ (-1)^{t-1} & \cdot & \cdot & \cdot & 0 & 0 \end{pmatrix}.$$

Then $XU = -U'X$, and as a consequence of (31) $H_{11} = XG$, where G is commutative with U. Hence

$$H_{11} = Xf(U), \tag{32}$$

where $f(U)$ is a polynomial in U, with coefficients in K.

Since $H_{11} = H'_{11}$,

$$Xf(U) = f(U')X' = X'f(-U).$$

Hence, if t is odd,

$$f(U) = f(-U) = g(U^2),$$

while, if t is even,

$$f(U) = -f(-U) = Ug(U^2).$$

Let t be odd. The congruent transformation by the matrix W_1 of (20) reduces a block of type (β) to one in which H_{11} is non-singular. Therefore we need only consider blocks of type (α). Let the matrix $f(U)$ in (32) be written as

$$f(u) = gE + \gamma, \text{ where } \gamma = \gamma(U^2) = U^2\gamma_1(U^2).$$

Then, if $W = E - \tfrac{1}{2}g^{-1}\gamma$,

$$\begin{aligned} W'Xf(U)NW &= X(E - \tfrac{1}{2}\gamma g^{-1})(gE + \gamma)(E - \tfrac{1}{2}g^{-1}\gamma)N, \\ &= X(gE + \gamma^2\phi)N, \end{aligned}$$

where ϕ is a polynomial in U^2. Since γ^2 contains a factor U^4, it is possible by a succession of such transformations to reduce H_{11} to the form XgE. The element g is not unique. By an argument similar to that of II, pages 152-154, we have

THEOREM 3. *If t is odd and r^t occurs c times among the elementary factors of the pencil Λ, in the canonical form of Λ occurs a block matrix $[g_1X_tN_t, g_2X_tN_t, \cdots, g_cX_tN_t]$, where the diagonal matrix $[g_1, g_2, \cdots, g_c]$ is determined apart from a congruent transformation.*

Let t be even. Then H_{11} is singular and only blocks of type (β) may occur. It is easily shown that

$$\begin{pmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{pmatrix} = \begin{pmatrix} XUf_{11}(U) & Xf_{12}(U) \\ Xf_{21}(U) & XUf_{22}(U) \end{pmatrix} = \gamma + \phi,$$

where $\gamma = \begin{pmatrix} 0 & Xf_{12}(U) \\ Xf_{21}(U) & 0 \end{pmatrix}$, $\gamma[N, N] = [N'', N'']\gamma'$ and

$$\phi[N, N] = [N'', N'']\phi'.$$

Therefore,

$$\begin{aligned}(E - \tfrac{1}{2}\phi\gamma^{-1})(\gamma + \phi)[N, N](E - \tfrac{1}{2}\phi\gamma^{-1})' &= (E - \tfrac{1}{2}\phi\gamma^{-1})(\gamma + \phi)\ (E - \tfrac{1}{2}\gamma^{-1}\phi)[N, N] \\ &= (\gamma - \phi\gamma^{-1}\phi + \tfrac{1}{4}\phi\gamma^{-1}\phi\gamma^{-1}\phi)[N, N], \\ &= (\gamma_1 + \phi_1)[N, N],\end{aligned}$$

where

$$\gamma_1 = \begin{pmatrix} 0 & Xg_{12}(U) \\ Xg_{21}(U) & 0 \end{pmatrix} \quad \text{and} \quad \phi_1 = \begin{pmatrix} XU^3g_{11}(U) & 0 \\ 0 & XU^3g_{22}(U) \end{pmatrix}.$$

Since γ is non-singular, γ_1 is non-singular, and we may repeat this process with γ replaced by γ_1 and ϕ by ϕ_1. After $t/2$ repetitions the matrix $\phi_{t/2}$ is the zero matrix, since $U^t = 0$. Hence a block of type (β) may be reduced to the form

$$\begin{pmatrix} 0 & H_{12} \\ H_{21} & 0 \end{pmatrix}\begin{pmatrix} N & 0 \\ 0 & N \end{pmatrix}, \tag{33}$$

where $H_{12}N = N''H'_{21}$. But,

$$\begin{aligned}\begin{pmatrix} E & 0 \\ 0 & H_{21}^{-1} \end{pmatrix}\begin{pmatrix} 0 & H_{12} \\ H_{21} & 0 \end{pmatrix}\begin{pmatrix} N & 0 \\ 0 & N \end{pmatrix}\begin{pmatrix} E & 0 \\ 0 & H_{21}^{-1} \end{pmatrix}' \\ = \begin{pmatrix} 0 & H_{12} \\ E & 0 \end{pmatrix}\begin{pmatrix} E & 0 \\ 0 & H_{12}^{-1} \end{pmatrix}\begin{pmatrix} N & 0 \\ 0 & N'' \end{pmatrix} = \begin{pmatrix} 0 & N'' \\ N & 0 \end{pmatrix}.\end{aligned} \tag{34}$$

Therefore we have proved,

THEOREM 4. *If t is even and r^t occurs c times among the elementary factors of Λ, then c is even. Corresponding to each pair of elementary factors r^t, r^t, there is a block matrix $\begin{pmatrix} 0 & N'' \\ N & 0 \end{pmatrix}$ in the canonical form of Λ.*

Theorems 1, 3 and 4 complete the solution of problem (b) when $\epsilon = 1$, $\delta = -1$.

Case 3. The matrices H_{11} and $H_{11}U$ are both skew symmetric so that $H_{11} = U'H_{11}$. If T is the counter unit matrix of order t,

$TU = U'T$ and $H_{11} = Tf(U)$. Since $H_{11} = -H'_{11}$,
$Tf(U) = -f(U)T = -Tf(U)$, so that $f(U) = 0$ and $H_{11} = 0$.

Therefore we need only consider blocks of type (β), $\begin{pmatrix} 0 & H_{12} \\ H_{21} & 0 \end{pmatrix}\begin{pmatrix} N & 0 \\ 0 & N \end{pmatrix}$

which is of the same nature as (33) and can be reduced to $\begin{pmatrix} 0 & N'' \\ N & 0 \end{pmatrix}$ by (34). Therefore,

THEOREM 5. *If r^t occurs c times among the elementary factors of a skew symmetric pencil Λ, c is even. Corresponding to each pair of elementary factors r^t, r^t there is a block matrix $\begin{pmatrix} 0 & N'' \\ N & 0 \end{pmatrix}$ in the canonical form of Λ.*

3. In sections 1 and 2 we have proved that there exists in all three cases a non-singular matrix P such that

$$P^*\Lambda P = [\Lambda_1, \Lambda_2]$$

where the form of Λ_1 is determined and

$$\Lambda_2 = H(rM + sE), \tag{35}$$

the matrix H being non-singular.

As mentioned in the introduction a normal form for Λ_2 has been determined in case 1 (I) and also in case 2 (II). We now consider case 3. The matrices H and HM are both skew symmetric, so that $HM = M'H$ and, consequently, if Q is any non-singular matrix satisfying the equation

$$QM = M'Q, \tag{36}$$

then $H = QG$, where $GM = MG$.

Let the elementary factors of $rM + sE$ be the homogeneous polynomials

$$p_i(r, s)^{\eta_{ij}}, \qquad (i = 1, 2, \cdots, t;\ j = 1, 2, \cdots, r_i), \tag{37}$$

where $p_i(r, s)$ is irreducible in $K[r, s]$.

We may take M in the canonical form

$$M = [M_1, M_2, \cdots, M_t],$$

where the elementary factors of $rM_k + sE_k$ are the polynomials (37), when $i = k$. Since there exists a non-singular matrix Q_k such that $Q_kM_k = M'_kQ_k$ (I page 478), the matrix $Q = [Q_1, Q_2, \cdots, Q_t]$ is non-singular and satisfies (36). Any matrix G commutative with M is also a diagonal block matrix $[G_1, G_2, \cdots, G_t]$ [11] and consequently

$$H = QG = [Q_1G_1, Q_2G_2, \cdots, Q_tG_t].$$

Therefore it is sufficient to consider a pencil Λ_2 in (35), whose elementary factors are all powers of the same irreducible polynomial $p(r, s)$ i. e. are

[11] John Williamson, "The idempotent and nilpotent elements of a matrix," *American Journal of Mathematics*, vol. 58 (1936), pp. 747-758.

$$p(r,s)^{\eta_i}, \qquad \eta_1 \geqq \eta_2 \geqq \cdots \geqq \eta_t.$$

Accordingly we take M in the canonical form

(38) $$M = [L_1, L_2, \cdots, L_t],$$

where

(39) $$L_i = p \cdot E_i + e \cdot U_i, \text{ and write}$$

(40) $$N_i = rL_i + se \cdot E_i.$$

In (39), E_i and U_i are respectively the unit matrix and the auxiliary unit matrix of order η_i; p is the companion matrix of $p(1,\lambda)$; e the unit matrix of the same order as p; and $\cdot$ denotes direct product. (I page 477). For example, if $\eta_t = 3$,

$$L_t = \begin{pmatrix} p & e & 0 \\ 0 & p & e \\ 0 & 0 & p \end{pmatrix}.$$

It has been shown, I page 490, that there exists a non-singular symmetric matrix q such that $qp = p'q$. Hence, if T_i is the counter unit matrix of order η_i,

$$q \cdot T_i L_i = q \cdot T_i(p \cdot E_i + e \cdot U_i) = (p' \cdot E_i + e \cdot U'_i) q \cdot T_i = L'_i q \cdot T_i.$$

Therefore the matrix

$$Q = [q \cdot T_1, q \cdot T_2, \cdots, q \cdot T_t]$$

satisfies (36), and is symmetric.

Let

$$G = (G_{ij}), \qquad (i, j = 1, 2, \cdots, t),$$

be a partition of a matrix G similar to that of M in (38). If G is commutative with M, the form of G is known.[12] In fact, if $\eta_i \geqq \eta_j$,

(41) $$G_{ij} = \begin{pmatrix} S_{ij} \\ 0 \end{pmatrix}, \qquad G_{ji} = (0, S_{ji}),$$

where S_{ij} and S_{ji} are square matrices of order $\eta = \eta_j$, while 0 is the zero matrix of $\eta_i - \eta_j$ rows and η_j columns. Further

(42) $$S_{ij} = \sum_{a=0}^{\eta-1} s_{ija} U_j^a, \qquad S_{ji} = \sum_{a=0}^{\eta-1} s_{jia} U_j^a,$$

where $s_{ijk} = s_{ijk}(p)$ is a polynomial in the matrix p.

Since $H = QG$ is skew symmetric,

$$QG = -G'Q' = -G'Q.$$

[12] John Williamson, *The Idempotent and Nilpotent Elements of a Matrix*, p. 457.

Hence $q \cdot T_i G_{ij} = - G'_{ji} q \cdot T_j$, or, when $i \leqq j$,

$$q \cdot T_i \cdot \begin{pmatrix} S_{ij} \\ 0 \end{pmatrix} = - \begin{pmatrix} 0 \\ S'_{ji} \end{pmatrix} q \cdot T_j.$$

From this we deduce

$$q \cdot T_j S_{ij} = - S'_{ji} q \cdot T_j.$$

But, by direct calculation from (42),

$$q \cdot T_j S_{ji} = S'_{ji} q \cdot T_j,$$

and hence

$$S_{ij} = - S_{ji}. \tag{43}$$

In particular $G_{ii} = S_{ii} = 0$. Let $\eta_1 = \eta_2 = \cdots = \eta_c > \eta_{c+1}$. Since G is non-singular and, when $j > c$, the first column of G_{j1} is zero, G_{i1} must be non-singular for at least one value of i, $2 \leqq i \leqq c$. We may therefore assume that G_{21} is non-singular and, as a consequence of (43), that

$$\begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} = \begin{pmatrix} 0 & G_{12} \\ G_{21} & 0 \end{pmatrix}$$

is non-singular. Therefore by Lemma 6 there exists a non-singular matrix P such that

$$P'\Lambda_2 P = \left[\begin{pmatrix} 0 & H_{12} \\ H_{21} & 0 \end{pmatrix} \begin{pmatrix} N_1 & 0 \\ 0 & N_1 \end{pmatrix}, H_2\Lambda_3 \right].$$

Since the block $\begin{pmatrix} 0 & H_{12} \\ H_{21} & 0 \end{pmatrix} \begin{pmatrix} N_1 & 0 \\ 0 & N_1 \end{pmatrix}$ is of the same type as (33), it may be reduced to the block matrix $\begin{pmatrix} 0 & N_1'' \\ N_1 & 0 \end{pmatrix}$.

We have therefore proved.

THEOREM 6. *Each elementary factor $p(r,s)^\eta$ of a pencil of skew symmetric matrices must occur an even number of times. Corresponding to each pair of elementary factors $p(r,s)^\eta$, $p(r,s)^\eta$, in the canonical form is a matrix block $\begin{pmatrix} 0 & -N' \\ N & 0 \end{pmatrix}$ where N is defined by (39) and (40).*

Combining this with the results of sections (1) and (2) we have finally,

THEOREM 7. *Necessary and sufficient conditions, that two pencils of skew symmetric matrices be equivalent under a non-singular congruent transformation in K, are that the two pencils have the same kronecker minimal row indices and the same elementary factors.*

THE JOHNS HOPKINS UNIVERSITY.

SOME REMARKS ON CLASS FIELD THEORY OVER INFINITE FIELDS OF ALGEBRAIC NUMBERS.*

By O. F. G. Schilling.

Mr. M. Moriya recently investigated the theory of finite abelian extensions over infinite fields of algebraic numbers.[1] He has shown that under certain restricting conditions on the infinite algebraic ground field there exists an analog to the classical class field theory: the finite abelian extensions of an infinite field can be characterized by class groups of ideals in the groundfield. His results can be completed in several directions. In this note we shall characterize the finite algebraic number fields by an intrinsic property of the given field: a number field is finite if and only if there exists a finite number of cyclic superfields of some prime degree with a given defining modulus. Furtherfore we shall prove the analog to the theorem on arithmetic progressions for a certain class of infinite fields. Finally we discuss the norm theorem of Hilbert and Hasse. This connects our investigation with A. A. Albert's results on algebras over infinite number fields.[2]

Let k be an infinite field of algebraic numbers over the field P of all rational numbers. The field k can always be approximated by an enumerable tower of finite algebraic number fields k_i over P; that is to say k is the join $\Sigma\, k_i$ of finite fields k_i such that

$$\cdots \subset k_{i-1} \subset k_i \subset \cdots \subset \Sigma\, k_i = k.$$

With the field k there is associated a Steinitz G-number $N(k, \mathrm{P})$, the absolute G-degree of k. The number $N(k, \mathrm{P})$ is defined as the formal least common multiple of all the relative degrees $[h:\mathrm{P}]$ where h is any finite subfield of k. If a prime p divides almost all degrees $[h:\mathrm{P}]$—or almost all degrees $[k_i:k_{i-1}]$ is sufficient too—then we say that p^∞ divides $N(k, \mathrm{P})$. Thus $N(k, \mathrm{P})$ can be uniquely decomposed into an infinite part $N_{\mathrm{inf}}(k, \mathrm{P})$ consisting of the product of all $p^\infty \mid N(k, \mathrm{P})$, and a finite part $N_{\mathrm{fin}}(k, \mathrm{P})$ which consists of the exact powers of those primes p which divide only a finite number of

* Received December 18, 1936.

[1] M. Moriya, "Klassenkörpertheorie für einen unendlichen Zahlkörper." Will appear in the *Journal of the Faculty of Science*, Sapporo (Japan).

[2] A. A. Albert, "Normal division algebras over algebraic number fields not of finite degree," *Bulletin of the American Mathematical Society*, October, 1933.

relative degrees $[h:\mathrm{P}]$. Moriya has shown that exactly all those abelian extensions K of k whose degrees $[K:k]=n$ are prime to $N_{\inf}(k,\mathrm{P})$ are class fields, that is to say the Galois groups $G(K,k)$ of K over k are isomorphic with class groups $\mathfrak{a}/(H(K,k)$ derived from the group $\mathfrak{a}$ of all ideals in k that have inverses.[3]

It is not difficult to construct infinite fields k whose infinite parts $N_{\inf}(k,\mathrm{P})$ of the respective G-degrees $N(k,\mathrm{P})$ are equal to one. Such fields k have then the property that all abelian fields K over k are class fields. There arises the problem of finding properties of these infinite fields k that distinguish them from the finite algebraic number fields.

Now let k_0 be a finite algebraic number field which contain the l-th roots of unity ζ_l $(l \neq 2)$. By $\mathfrak{f}$ we denote an integral ideal of k whose prime divisors we shall later specify, and by ∞ the product over all the infinite prime places of k_0.[4] Assume that $\mathfrak{f}$ is chosen in such a fashion that there exist cyclic extensions Z of k_0 whose degrees are equal to l and whose conductors $\mathfrak{f}(Z,k_0)$ are divisors of $\mathfrak{f}\infty$. Such moduli $\mathfrak{f}\infty$ always exist. The number of different cyclic fields Z of this type shall be denoted by $R(k_0,l,\mathfrak{f}\infty)$.

Let the prime divisors of l in k_0 be $\mathfrak{l}_1,\cdots,\mathfrak{l}_s$; we obtain a decomposition $(l)=\prod_{i=1}^{s}\mathfrak{l}_i^{e(i)}$, and have $N(k_0,\mathrm{P}|\mathfrak{l}_i)=(l)^{f(i)}$ where $\sum_{i=1}^{s}e(i)f(i)=[k_0:P]$. Then the numbers $w(k_0,\mathfrak{l}_i)=e(i)l(l-1)^{-1}$ are always integral.[5] Suppose now that the ideal $\mathfrak{f}=\prod_{(\mathfrak{p},\mathfrak{l})=1}\mathfrak{p}\prod_{\mathfrak{l}_i|l}\mathfrak{l}_i^{v(i)}$ is chosen in such a way that the exponents $v(i)$ are greater than or equal to the numbers $w(k_0,\mathfrak{l}_i)$ and that there occur sufficiently many prime ideals $\mathfrak{p}$. Then there surely exist cyclic fields Z of degree l over k whose conductors $\mathfrak{f}(Z,k)$ divide $\mathfrak{f}\infty$.

LEMMA 1. *If k'_0 is a finite extension of k_0 whose relative degree n' is prime to l then*

$$R(k'_0,l,\mathfrak{f}\infty) > R(k_0,l,\mathfrak{f}\infty).$$

Proof. First we give an explicit expression for the number $R(k_0,l,\mathfrak{f}\infty)$. By $w(\mathfrak{p})$ and $w(\mathfrak{l})$ we denote the number of prime divisors $\mathfrak{p}$ and $\mathfrak{l}$ which divide $\mathfrak{f}$. The number of fundamental basic units of k_0 is equal to $r(k_0)=[k_0:\mathrm{P}]\,2^{-1}-1$ because there exist no real infinite prime places in k_0 and the number of complex infinite prime places is equal to $[k_0:\mathrm{P}]\,2^{-1}$ as the following inclusion shows

[3] For details see the Paper of M. Moriya mentioned under no. 1.

[4] For the class field theory over finite number fields see the Tract of H. Hasse in vol. 35 (1930) of the *Jahresberichte der Deutschen Mathematikervereinigung.*

[5] The following formulae can be found in Hasse's Tract, part Ia, § 15.

[6] Cf. Hasse's Tract, part II.

$$\mathrm{P} \subset \mathrm{P}(\zeta_l + \zeta_l^{-1}) \subset \mathrm{P}(\zeta_l) \subseteq k.$$

Now let $\{\alpha\}$ be the multiplicative group of all numbers α in k_0 which are prime to $\mathfrak{f}$. The group $\{\alpha\}$ contains the subgroup $\{\omega\}$ of all numbers ω in k_0 for which $\omega \equiv \alpha^l \pmod{\mathfrak{f}}$ and $(\omega) = \mathfrak{r}^l$ where $\mathfrak{r}$ are non-principal ideals of k_0 which are relatively prime to $\mathfrak{f}$. (Written as $\mathfrak{r} \not\sim 1(k_0)$.) Then the index $[\{\alpha\}:\{\omega\}]$ is equal to the $m(k_0)$-th power of l. Finally $R(k_0, l, \mathfrak{f}\infty)$ becomes equal to $(l^S - 1)(l-1)^{-1}$ where

$$\begin{aligned} S &= w(\mathfrak{p}) + w(\mathfrak{l}) + m(k_0) + \sum_{i=1}^{s} e(i) f(i) - (r(k_0) - 1) \\ &= w(\mathfrak{p}) + w(\mathfrak{l}) + m(k_0) + [k_0 : P]\, 2^{-1}. \end{aligned}$$

Now we wish to calculate the expression S' in k' with respect to the same modulus $\mathfrak{f}\infty$. The prime ideal $\mathfrak{p}$ may be decomposed as $\mathfrak{p} = \prod_{j=1}^{g(\ldots)} \mathfrak{p}'_j{}^{e(j)}$ where $g(\ldots) = g(\mathfrak{p})$; then $w(\mathfrak{p}') \geq w(\mathfrak{p})$ and the exponents $e(j)$ can be normalized to one because they do not matter in the determination of S'. For the prime ideals $\mathfrak{l}_i$ we obtain similarly $\mathfrak{l}_i = \prod_{j=1}^{g(i)} \mathfrak{l}'^{e(i,j)}_{ij}$, and $\Pi\, \mathfrak{l}_i^{v(i)}$ goes over into $\prod_{i,j} \mathfrak{l}'^{v(i)e(i,j)}_{ij}$. As a simple calculation shows, we have

$$v(i) e(i,j) \geq w(k'_0, \mathfrak{l}'_{ij}) = e(i,j) w(k_0, \mathfrak{l}_i);$$

hence $w(\mathfrak{l}') \geq w(\mathfrak{l})$.

Moreover, $m(k'_0) \geq m(k_0)$. This is seen as follows. Evidently it is sufficient to show that an ideal $\mathfrak{r} \not\sim 1\ (k_0)$ cannot become a principal ideal in k'. Assume that $\mathfrak{r} \sim 1\ (k'_0)$, then $\boldsymbol{N}(k', k | \mathfrak{r}) = \mathfrak{r}^{n'} \sim 1\ (k_0)$. According to our assumptions, $(n', l) = 1$; hence there exist two integers c and d such that $cn' + dl = 1$. This leads to the relation $\mathfrak{r}^1 = \mathfrak{r}^{cn'+dl} = (\mathfrak{r}^{n'})^c (\mathfrak{r}^l)^d \sim 1\ (k_0)$, for $\mathfrak{r}^l \sim 1\ (k_0)$. But this is in contradiction to the assumptions on the ideal $\mathfrak{r}$. Thus we obtain

$$\begin{aligned} S' &= w(\mathfrak{p}') + w(\mathfrak{l}') + m(k'_0) + [k' : \mathrm{P}]\, 2^{-1} \\ &= w(\mathfrak{p}') + w(\mathfrak{l}') + m(k'_0) + n'[k_0 : \mathrm{P}]\, 2^{-1}, \end{aligned}$$

so that certainly $S' > S$ for $n' > 1$. Thus we have

$$R(k'_0, l, \mathfrak{f}\infty) > R(k_0, l, \mathfrak{f}\infty).$$

Now let k be an infinite algebraic extension of k_0 such that $N_{\inf}(k, \mathrm{P})$ is prime to l.

Lemma 2. *The number of cyclic extensions $\mathfrak{Z}$ of degree l over k whose conductors are divisors of $\mathfrak{f}\infty$ is infinite.*

Proof. Each algebraic extension $K = k(\vartheta)$ of k is evidently the join of k and a finite field K_0. For take some finite subfield k_0 of k such that the coefficients belonging to the irreducible equation of ϑ in k lie in it; then $k_0(\vartheta)$ is a field with the asserted property. Thus all cyclic fields Z of degree l are joins of finite cyclic fields Z_j over $k_j \subset k$ with k. Now $R(k_i, l, \mathfrak{f}\infty) > R(k_{i-1}, l, \mathfrak{f}\infty)$ for sufficiently high i, because $([k_i : k_{i-1}], l) = 1$ as a consequence of $(N_{\inf}(k, \mathrm{P}), l) = 1$. Thus there exist infinitely many cyclic fields Z with the described property.

THEOREM 1. *If k is an algebraic number field such that for a prime $l \neq 2$ the class field theory holds, then k is infinite if and only if there exist infinitely many cyclic superfields Z of degree l over k whose conductors are divisors of a given modulus $\mathfrak{f}\infty$ lying in a finite subfield of k.*

Proof. Without loss of generality we can assume that k contains the l-th roots of unity, because $k(\zeta_l)$ is at most of degree $l - 1$ over k. Therefore the assumptions on k are carried over to $k(\zeta_l)$, $N_{\inf}(k, \mathrm{P}) = N_{\inf}(k(\zeta_l), \mathrm{P})$.

Now if k is infinite we denote by k_0 some finite subfield of k such that it contains the l-th roots of unity and that $N(k, k_0)$ is prime to l. The modulus $\mathfrak{f}\infty$ shall be chosen as a modulus of the type investigated before; then from the fact that for l the class field theory holds—that is to say, all cyclic fields of degree l over k can be described uniquely by class groups of ideals in k—it follows that $(N_{\inf}(k, \mathrm{P}), l) = 1$. Hence Lemma 2 can be applied on k, it asserts that there exist infinitely many fields Z whose conductors divide $\mathfrak{f}\infty$. Conversely, if there exist infinitely many fields Z with these properties then k is infinite. This is obvious because a finite field never possesses an infinity of cyclic superfields of degree whose conductors divide a fixed modulus $\mathfrak{f}\infty$.

THEOREM 2. *If K is a class field of degree n over the infinite algebraic number field k then there exists for every divisor f of n, where f is the order of an element of the Galois group $G(K, k)$, an infinity of prime ideals $\mathfrak{p}$ in k which are prime to the discriminant of K over k and whose prime divisors $\mathfrak{P}$ in K have the relative residue class degree f.*

Proof. We chose a finite subfield k_0 in k such that $N(k, k_0)$ is prime to n; this is always possible because K was assumed to be a class field of degree n over k. If $K = k(\vartheta)$ then there exists some finite field k_i in k which contains k_0 such that $K = K_i k$ where $K_i = k_i(\vartheta)$. According to the theorem on arithmetic progressions in k_i there exist infinitely many prime ideals $\mathfrak{p}(i)$ in k which are prime to the discriminant of K_i over k_i and whose divisors $\mathfrak{P}(i, j(i))$ in K_i are of relative residue degrees f over k_i, if f is the order of an arbitrary

element S_i of the Galois group $G(K_i, k_i) \cong G(K, k)$. Now the Artin symbol $(K_i, k_i/\mathfrak{p}(i)) = S_i$. Let $k_{i+\nu}$ be any finite extension of k_i which belongs to the approximation $\{k_i\}$ of k, and let $\mathfrak{p}(i+\nu)$ be any divisor of a fixed $\mathfrak{p}(i)$. The translation theorem of class field theory asserts that $(K_{i+\nu}, k_{i+\nu}/\mathfrak{p}(i+\nu))$ is also of order f because the degree $[k_{i+\nu} : k_i]$ is prime to n. The residue class degree of any divisor $\mathfrak{P}(i+\nu, j(i+\nu))$ is therefore equal to f for any ν. Now take a sequence $\{\mathfrak{p}(i+\nu)\}$ of prime ideals $\mathfrak{p}(i+\nu)$ such that $\mathfrak{p}(i) \subseteq \mathfrak{p}(i+1) \subseteq \cdots \subseteq \mathfrak{p}(i+\nu) \subseteq \cdots$. This sequence determines a prime ideal $\mathfrak{p}$ in k and it is prime to the discriminant of K with respect to k. In the same fashion we determine a chain of prime ideals $\mathfrak{P}(i+\nu, j(i+\nu))$ in $K_{i+\nu}$ such that

$$\text{i)} \quad \mathfrak{P}(i, j(i)) \subseteq \mathfrak{P}(i+1, j(i+1)) \subseteq \cdots \subseteq \mathfrak{P}(i+\nu, j(i+\nu)) \subseteq \cdots$$

$$\text{and ii)} \quad \mathfrak{p}(i+\nu) \subseteq \mathfrak{P}(i+\nu, j(i+\nu)).$$

The limit prime ideal $\mathfrak{P}$ of $\{\mathfrak{P}(i+\nu, j(i+\nu))\}$ in K is then a divisor of $\mathfrak{p}$, and the residue class degree of $\mathfrak{P}$ is equal to f. The equalities

$$f = f(i) = f(i+1) = \cdots f(i+\nu) = \cdots$$

for the respective residue class degrees assert according to Herbrand the norm relation.[7]

Now the number of prime ideals $\mathfrak{p}(i)$ belonging to a fixed order f is infinite; hence the proof of the theorem is complete.

Now let $\mathfrak{p}$ be an arbitrary prime ideal of the infinite algebraic number field k. Then $\mathfrak{p}$ uniquely determines a valuation on the field k. The system of all fundamental sequences $\{\alpha_\mu\}$—α_μ in k—with respect to that valuation form a field $k(\mathfrak{p})$, the so-called derived field of k with respect to $\mathfrak{p}$. If k is equal to $\Sigma\, k_i$, then the intersections $\mathfrak{p}(i) = \mathfrak{p} \cap k_i$ are prime ideals in the finite subfields k_i of k. Form $\Sigma\, k_i(\mathfrak{p}(i))$. This field is in general an infinite algebraic extension of $k_0(\mathfrak{p}(0))$ and it is not closed with respect to $\mathfrak{p}$; but the following lemma holds.

Lemma 3. *The derived field $k(\mathfrak{p})$ of an infinite algebraic number field $k = \Sigma\, k_i$ is equal to the derived field belonging to the field $\Sigma\, k_i(\mathfrak{p}(i))$ where $k_i(\mathfrak{p}(i))$ denotes the perfect fields of k_i with respect to $\mathfrak{p}(i) = \mathfrak{p} \cap k_i$.*

Proof. According to the construction of the valuation belonging to $\mathfrak{p}$ in k the value groups of $\mathfrak{p}$ in k and of $\mathfrak{p}' = \mathfrak{p} \cap \Sigma\, k_i(\mathfrak{p}(i))$ coincide.[8]

[7] J. Herbrand, "Théorie arithmétique des corps de nombres de degré infini, I. Extensions de degré fini," *Mathematische Annalen*, vol. 106 (1932).

[8] W. Krull, "Idealtheorie in unendlichen Zahlkörpern," *Mathematische Zeitschrift*, vol. 29 (1928).

First we show that $k(\mathfrak{p})$ is contained in the derived field of $\Sigma\, k_i(\mathfrak{p}(i))$. Let $\{\alpha_\mu\}$ be an arbitrary fundamental sequence of elements α_μ in k, it represents an arbitrary element of $k(\mathfrak{p})$. Each of the elements α_μ lies already in a suitable finite subfield of k. Therefore the sequence $\{\alpha_\mu\}$ consists of elements in $\Sigma\, k_i(\mathfrak{p}(i))$; it is also a fundamental sequence with respect to $\mathfrak{p}'$ because the value groups of $\Sigma\, k_i(\mathfrak{p}(i))$ and k resp. $k(\mathfrak{p})$ coincide. Hence $\{\alpha_\mu\}$ lies in the derived field of $\Sigma\, k_i(\mathfrak{p}(i))$.

Conversely, each fundamental sequence $\{\alpha'_\mu\}$ of $\Sigma\, k_i(\mathfrak{p}(i))$ is an element of $k(\mathfrak{p})$. According to the definition of $\Sigma\, k_i(\mathfrak{p}(i))$ each element α'_μ lies already in a finite $\mathfrak{p}(\mu)$-adic field $k_\mu(\mathfrak{p}(\mu)) \supseteq k_\mu$; therefore we always can choose some element α_μ in k which lies arbitrarily near to a'_μ in the sense of the valuation belonging to $\mathfrak{p}'$. The sequence $\{\alpha_\mu\}$ is also a fundamental sequence of $\Sigma\, k_i(\mathfrak{p}(i))$, and according to the definition of the closure of $\Sigma\, k_i(\mathfrak{p}(i))$ we have $\{\alpha_\mu\} = \{\alpha'_\mu\}$. The sequence $\{\alpha_\mu\}$ is a fundamental sequence of k, therefore $\{\alpha'_\mu\} = \{\alpha_\mu\}$ lies in $k(\mathfrak{p})$.

M. Moriya and the author have proved in 2 papers that the class field theories over $\Sigma\, k_i(\mathfrak{p}(i))$ and $k(\mathfrak{p})$ are virtually the same. There exists a one-to-one correspondence between the abelian finite extensions and finite normal algebras over both fields respectively.[9] Therefore it is of no importance in which field arithmetic investigations are made. For convenience we shall work in $\Sigma\, k_i(\mathfrak{p}(i))$ in the following considerations.

THEOREM 3. *If Z is a cyclic class field of degree n over the infinite algebraic number field k, then an algebra $A = (a, Z/k)$ is a complete matric algebra if and only if $A(\mathfrak{p}) = (a, Z/k) \times k(\mathfrak{p})$ are complete matric algebras for all prime divisors $\mathfrak{p}$ of k.*

Proof. According to what we just stated the relation $(a,Z/k) \times k(\mathfrak{p}) \sim k(\mathfrak{p})$ is equivalent to $(a, Z/k) \times \Sigma\, k_i(\mathfrak{p}(i)) \sim \Sigma\, k_i(\mathfrak{p}(i))$. Now let $k_0 \subset k$ be a finite subfield such that $N(k, k_0)$ is prime to n; such a field k_0 always exists if Z is a class field of degree n. Now $Z = k(\vartheta)$; let us take an extension k_* of k_0 such that a and the coefficients belonging to the irreducible equation of ϑ in k lie in k_*, write $Z_* = k_*(\vartheta)$. Hence $(a, Z/k) \sim (a, Z_*/k_*) \times k$.

Now assume that $(a, Z/k) \not\sim k$ although

$$(a, Z/k) \times \Sigma\, k_i(\mathfrak{p}(i)) \sim \Sigma\, k_i(\mathfrak{p}(i))$$

for all $\mathfrak{p}$ in k, the infinite prime spots included. Then also $(a, Z_*/k_*) \not\sim k_*$.

[9] M. Moriya and O. F. G. Schilling, "Zur Klassenkörpertheorie über unendlichen perfekten Körpern," and an additional note. Both to appear in the forthcoming volume of the Sapporo Journal.

According to the fundamental theorem on normal algebras over finite number fields there exists at least one prime *ideal* $\mathfrak{p}_*$ in k_* such that

$$(a, Z_*/k_*) \times k_*(\mathfrak{p}_*) \not\sim k_*(\mathfrak{p}_*).^{10}$$

The field $k_*(\mathfrak{p}_*)$ is a subfield of $\Sigma\, k_i(\mathfrak{p}(i))$. Our assumptions yield

$$(a, Z_*/k_*) \times k_*(\mathfrak{p}_*) \times \Sigma\, k_i(\mathfrak{p}(i)) \sim (a, Z_*/k_*) \times \Sigma\, k_i(\mathfrak{p}(i)) \sim \Sigma\, k_i(\mathfrak{p}(i)),$$

that is to say that $\Sigma\, k_i(\mathfrak{p}(i))$ is a splitting field of $(a, Z_*/k_*) \times k_*(\mathfrak{p}_*)$. The splitting must already be finished in a finite extension $k_\lambda(\mathfrak{p}(\lambda))$ of $k_*(\mathfrak{p}_*)$ because $\Sigma\, k_i(\mathfrak{p}(i))$ is algebraic over $k_*(\mathfrak{p}_*)$. Hence according to the local class field theory over finite $\mathfrak{p}_*$-adic fields, the degree $[k_\lambda(\mathfrak{p}(\lambda)) : k_*(\mathfrak{p}_*)]$ must be a multiple of the exponent belonging to the algebra $(a, Z_*/k_*) \times k_*(\mathfrak{p}_*)$. The latter is a divisor of n and certainly different from one if

$$(a, Z_*/k_*) \times k_*(\mathfrak{p}_*) \not\sim k_*(\mathfrak{p}_*).$$

Hence $([k_\lambda(\mathfrak{p}(\lambda)) : k_*(\mathfrak{p}_*)], n) \neq 1$ and a fortiori $([k_\lambda : k_*], n) \neq 1$. But this contradicts the choice of k_* in k. Therefore we must have

$$(a, Z_*/k_*) \times k_*(\mathfrak{p}_*) \sim k_*(\mathfrak{p}_*)$$

and hence $\mathfrak{p}_*$ cannot be a ramified prime ideal. This means that $(a, Z_*/k_*) \sim k_*$ and a fortiori that $(a, Z/k) \sim (a, Z_*/k_*) \times k \sim k$.—As always the converse is trivial.

Remark. Theorem 3 holds of course for arbitrary simple algebras A over the field k because they all possess cyclic representations.

Now let $k = \Sigma\, k_i$ be an arbitrary infinite algebraic number field. We assume that there exists an algebra A of degree n over k which is not isomorphic with a complete matric algebra over k. The algebra A then can be represented in the form $A_0 \times k$ where A_0 is an algebra A_* over a suitably chosen finite subfield k_* of k. Then we have also $A_* \not\sim k_*$ and therefore there exists at least one prime ideal $\mathfrak{p}(0) = \mathfrak{p}_*$ in the field $k_0 = k_*$ which we shall take as field to start with the approximation $\{k_i\}$ of k, such that

$$A_0 \times k_0(\mathfrak{p}(0)) \not\sim k_0(\mathfrak{p}(0)).$$

The exponent of $A_0 \times k_0(\mathfrak{p}(0))$ may be denoted by $m(\mathfrak{p}_*)$. By $\mathfrak{p}$ we denote any prime divisor of $\mathfrak{p}_*$ in k; then $\mathfrak{p} = \lim \mathfrak{p}(i)$ where

$$\mathfrak{p}_* = \mathfrak{p}(0) \subseteq \mathfrak{p}(1) \subseteq \cdots \subseteq \mathfrak{p}(i) \subseteq \cdots.$$

[10] For the theory of normal algebras see H. Hasse, "Über die Struktur etc., . . . ," *Mathematische Annalen*, vol. 107 (1933).

And the G-degree

$$[\Sigma\, k_i(\mathfrak{p}(i)) : k(\mathfrak{p}(0))] = N(k, k_0; \mathfrak{p}) = N(k, k_*; \mathfrak{p})$$

is equal to the least common multiple of all the degrees

$$[k_i(\mathfrak{p}(i)) : k_{i-1}(\mathfrak{p}(i-1))].$$

LEMMA 4. *If $A = A_* \times k$ is a proper non-matric algebra of degree n over the infinite field $k = \Sigma\, k_i$ then there exists at least one prime ideal $\mathfrak{p}$ in k lying over a prime ideal $\mathfrak{p}_*$ of $k_* \subset k$, such that*

$$N(k, k_*; \mathfrak{p}) \not\equiv 0 \pmod{m(\mathfrak{p}_*)}.$$

Proof. The relation $A = A_* \times k \not\sim k$ asserts that k is not a splitting field of A_*. Therefore no finite extension k_i of $k_* = k_0$ is a splitting field of A_*. Hence there exists according to Hasse's criterion on finite splitting fields at least one prime ideal $\mathfrak{p}(0)$ in k_0 such that

$$[k_i(\mathfrak{p}(i)) : k_0(\mathfrak{p}(0))] \not\equiv 0 \pmod{m(\mathfrak{p}(0))}$$

if $\mathfrak{p}(i)$ is any prime divisor of $\mathfrak{p}(0) = \mathfrak{p}_*$ in k_i. If we now select a sequence

$$\mathfrak{p}_* = \mathfrak{p}(0) \subseteq \mathfrak{p}(1) \subseteq \cdots \subseteq \mathfrak{p}(i) \subseteq \cdots$$

then its limit prime ideal $\mathfrak{p}$ has the property that

$$N(k, k_*; \mathfrak{p}) \not\equiv 0 \pmod{m(\mathfrak{p}_*)}.$$

Obviously there exist in general many prime ideals $\mathfrak{p}$ lying over $\mathfrak{p}_*$ for which this relation is fulfilled.

Now we are able to extend Theorem 3 to arbitrary fields k.

THEOREM 4. *Any normal algebra A over k of finite degree is a matric algebra over k if and only if $A \times k(\mathfrak{p}) \sim k(\mathfrak{p})$ for all prime divisors $\mathfrak{p}$ of k.*

Proof. Assume that $A \not\sim k$, in spite of $A \times k(\mathfrak{p}) \not\sim k(\mathfrak{p})$, for all $\mathfrak{p}$. According to Lemma 4 there would exist a finite subfield k_* of k and prime ideals $\mathfrak{p}_*$ and $\mathfrak{p}$ such that $N(k, k_*; \mathfrak{p}) \not\equiv 0 \pmod{m(\mathfrak{p}_*)}$. This contradicts the assumption $A \times k(\mathfrak{p}) \sim k(\mathfrak{p})$ which is equivalent to

$$A \times \Sigma\, k_i(\mathfrak{p}(i)) \sim \Sigma\, k_i(\mathfrak{p}(i)),$$

for the latter asserts that

$$[k_j(\mathfrak{p}(j)) : k_*(\mathfrak{p}_*)] \equiv 0 \pmod{m(\mathfrak{p}_*)}$$

for suitable j.

Finally we wish to show by an example that there exist proper division algebras D of degree n over certain infinite algebraic number fields k although the G-degree of k is divisible by n^∞.

Let k_0 be an arbitrary algebraic number field, and let n be an arbitrary positive integer. Suppose that $\mathfrak{p}(0)^\nu$ $(\nu = 1, 2, \cdots, s)$ is any finite set of prime divisors in k containing at least one prime ideal and such that we can attribute to each $\mathfrak{p}(0)^\nu$ a rational fraction of maximal denominator n $\rho(0)^\nu$ mod 1 for which $\sum_\nu \rho(0)^\nu \equiv 0 \pmod 1$, where one of them has exactly the denominator n. Then there exists a uniquely determined division algebra D_0 over k_0 in which the $\mathfrak{p}(0)^\nu$ are ramified and which has exactly the exponent n. We now proceed to construct an infinite algebraic extension k of k_0 such that $D_0 \times k$ is still a division algebra and such that $N(k, k_0) = n^\infty$. According to a theorem of Grunwald there exist infinitely many abelian fields k^i of degree n over k such that the prime divisors $\mathfrak{p}(0)^\nu$ are totally decomposed in each of them.[11] According to the criterion on splitting fields none of the fields k^i is a splitting field of D. The fields

$$k_0,\ k_0k^1 = k_1, = k_1,\ k_1k^2 = k_2, \cdots, k_{i-1}k^i = k_i, \cdots$$

define an infinite field k of G-degree n over k_0. The field k is obviously not a splitting field of D, because no finite subfield of k is splitting field. For the same reason D remains a division algebra. We may mention that

$$k(\mathfrak{p}^\nu) \cong k_0(\mathfrak{p}(0)^\nu)$$

for a prime ideal $\mathfrak{p}^\nu = \lim \mathfrak{p}(i)^\nu$, $\mathfrak{p}(i)^\nu \mid \mathfrak{p}(0)^\nu$ in k_i.

THE INSTITUTE FOR ADVANCED STUDY,
PRINCETON, N. J.

[11] W. Grundwald, "Ein allgemeines Existenztheorem für algebraische Zahlkörper," *Journal für Mathematik*, vol. 169 (1933).

ON THE ADDITION OF CONVEX CURVES. II.*

By Richard Kershner.

By the vectorial sum $C_1(+)C_2$ of two convex curves C_1 and C_2 is meant the set of all points which may be represented in at least one way as the vectorial sum of a point on C_1 and a point on C_2. It has been shown by Bohr [1] that $C_1(+)C_2$ is either the closed interior of a convex curve C_E or is the closed annular region between two convex curves C_E and C_I, where C_I lies wholly within C_E. The outer boundary C_E of $C_1(+)C_2$ was discussed by Haviland,[2] who found very precise relationships between C_E and the component curves C_1, C_2 by the use of the Minkowski supporting function. For example, if C_1 and C_2 each possess a continuous positive radius of curvature then so does C_E and, in fact, if $\rho_1(\theta), \rho_2(\theta), \rho_E(\theta)$ are the radii of curvature of C_1, C_2, C_E respectively, at the point of C_1, C_2, C_E where the oriented normal has the inclination θ then $\rho_E(\theta) = \rho_1(\theta) + \rho_2(\theta)$.

Recently the author [3] has investigated the inner boundary curve C_I of $C_1(+)C_2$. The results obtained are similar to those of Haviland but the methods are essentially more complicated. The great distinction between the treatment and the results in the two cases is illustrated by the fact that the curve C_I may possess corners while both C_1 and C_2 are analytic.[4] This fact contrasts remarkably with the above remarks concerning the radius of curvature $\rho_E(\theta)$ of C_E.

The purpose of the present note is a discussion of the possible existence of corners in the curve C_I. Specifically, it will be shown that if C_1 and C_2 are analytic curves, then C_I can have but a finite number of corners. (It will be shown that this number may be arbitrarily large). On the other hand, if it be only required of C_1 and C_2 that they possess radii of curvature which

---

* Received February 28, 1937.

[1] H. Bohr, "Om Addition of uendelig mange konvekse Kurver," *Danske Videnskabernes Selskab* (Forhandlinger, 1913), pp. 325-366. For a short presentation of the proof of this fact cf. B. Jessen and A. Wintner, "Distribution functions and the Riemann zeta function," *Transactions of the American Mathematical Society*, vol. 38 (1935), p. 69.

[2] E. K. Haviland, "On the addition of convex curves in Bohr's theory of Dirichlet series," *American Journal of Mathematics*, vol. 55 (1933), pp. 332-334.

[3] R. Kershner, "On the addition of convex curves," *American Journal of Mathematics*, vol. 58 (1936), pp. 737-746.

[4] Cf., e. g., R. Kershner, "On the values of the Riemann ζ-function on fixed lines $\sigma > 1$," *American Journal of Mathematics*, vol. 59 (1937), pp. 167-174.

can be differentiated infinitely often, then it is possible that C_I have an infinite number of corners. Finally it will be shown that if C_1 and C_2 have each a continuous radius of curvature then the corners of C_I are nowhere dense.

In the sequel it will always be assumed that C_I exists. Then [5] one of the curves C_1, C_2 may be placed in the other, after a rotation through the angle π about the origin, by a translation. It will be assumed that C_1 is the "larger" of the two curves so that C_2 may be placed in C_1, in the manner indicated above. By a point of C_1, C_2, C_I, in the direction θ, or, briefly, a point θ, is meant a point where the oriented normal has the inclination θ. Every point of C_I, except a corner, has a direction in the cases to be considered. A corner of C_I will be said to have the direction (θ_1, θ_2) if θ_1 and θ_2 are respectively the lower and upper limits of the directions of points in the neighborhood of the corner.

Using these notations we prove

LEMMA I. *If C_I has a corner in the direction (θ_1, θ_2) and if C_1 and C_2 have continuous, positive radii of curvature $\rho_1(\theta)$ and $\rho_2(\theta)$ respectively, then $\rho_1(\theta) = \rho_2(\theta + \pi)$ for some θ in $\theta_1 < \theta < \theta_2$; and, on the other hand, if $\rho_1(\theta) \leqq \rho_2(\theta + \pi)$ for some interval $\theta'_1 < \theta < \theta'_2$ then C_I has a corner in the direction (θ_1, θ_2) where $\theta_1 \leqq \theta'_1 < \theta'_2 \leqq \theta_2$.*

Proof. In terms of the mechanical interpretation [6] of C_I it is clear that the existence of a corner of C_I in the direction (θ_1, θ_2) means that the curve C_2, after being rotated through an angle π about the origin, may be placed within C_1 in such a way that it has internal contact with C_1 at the two points of C_1 which have the directions θ_1, θ_2. Lemma I follows [7] immediately from this fact. Lemma I gives immediately

THEOREM I. *If C_1 and C_2 are analytic curves then C_I has at most a finite number of corners.*

For suppose C_I had an infinite number of corners. Then the function $\rho_1(\theta) - \rho_2(\theta + \pi)$, where $\rho_i(\theta)$ is the radius of curvature of C_i, would have zeros clustering as some particular point θ. But since $\rho_i(\theta)$ is regular analytic this would imply $\rho_1(\theta) \equiv \rho_2(\theta + \pi)$ and C_I would not exist.

Now if $\rho(\theta)$ is a positive, continuous, periodic function of θ, with period

---

[5] R. Kershner, *loc. cit.* 3, Theorem I_0, p. 738.

[6] R. Kershner, *loc. cit.* 3, p. 741.

[7] Cf., e. g., S. Mukhopadhyaya, "Circles incident on an oval of undefined curvature," *Tôhoku Mathematical Journal*, vol. 34 (1931), pp. 115-129.

2π, then there will exist a closed convex curve of which $\rho(\theta)$ is the radius of curvature, if and only if the closure conditions

$$(1) \qquad \int_0^{2\pi} \rho(\theta) \cos\theta d\theta = 0; \qquad \int_0^{2\pi} \rho(\theta) \sin\theta d\theta = 0$$

are satisfied.[8] Using this fact it is very easy to show that the number of corners of C_I may be arbitrarily large even when both C_1 and C_2 are analytic. For let C_2 be a circle of radius r. Let

$$(2) \qquad \rho_1^{(n)}(\theta) = r + 1 + \cos n\theta - \delta_n, \qquad (n = 2, 3, 4, \cdots)$$

where $\delta_n > 0$ is chosen so small that, first of all, $\rho_1^{(n)}(\theta)$ is everywhere positive, so that ((1) being obviously satisfied) there does exist a closed convex curve $C_1^{(n)}$ having $\rho_1^{(n)}(\theta)$ as radius of curvature; and secondly that C_2 may be placed entirely within $C_1^{(n)}$. In order to satisfy the first requirement on δ_n it is obviously enough to choose $\delta_n < r$. To see that the second requirement may be satisfied it is enough to notice that if $\delta_n = 0$ then the corresponding $C_1^{(n)}$ has a radius of curvature never less and sometimes greater than r so that C_2 can be placed entirely within $C_1^{(n)}$ by a known theorem.[9] Now let δ_n be fixed satisfying the above requirements and consider the corresponding $C_1^{(n)}$. There are clearly n disjoint θ-intervals in which the radius of curvature (2) of $C_1^{(n)}$ is less than r. Then, by Lemma I, the inner boundary $C_I^{(n)}$ of the vectorial sum $C_1^{(n)}(+)C_2$ will have a corner in the directions (θ_i, θ'_i) for a set of direction intervals including these n disjoint θ-intervals. In general it is not true that two disjoint intervals in which $\rho_1^{(n)}(\theta) \leqq \rho_2(\theta + \pi)$ correspond to distinct corners but in this case the symmetry of $C_1^{(n)}$ makes it obvious that the n intervals mentioned above actually correspond to n distinct corners. Thus, *the inner boundary $C_I^{(n)}$ of the analytic convex sum $C_1^{(n)}(+)C_2$, where $C_1^{(n)}$ is defined by its radius of curvature* (2) *and C_2 is a circle of radius r, has n corners.*

THEOREM II. *There exist convex curves C_1, C_2 whose radii of curvature $\rho_1(\theta)$, $\rho_2(\theta)$ have infinitely many derivatives and which are such that the vector sum $C_1(+)C_2$ has an inner boundary C_I with infinitely many corners.*

Proof. Let C_2 be again a circle of radius r so that $\rho_2(\theta) \equiv r$. Let $\theta_0, \theta_1, \cdots \theta_n, \cdots$ be an infinite sequence of θ-values such that

$$(3) \qquad \pi/2 = \theta_0 > \theta_1 > \cdots > \theta_n \to 0.$$

Let

$$(4) \qquad \rho_1(\theta) \equiv r \text{ if } \theta_{2n} \geqq \theta \geqq \theta_{2n+1}; \qquad (n = 0, 1, 2, \cdots)$$

[8] W. Blaschke, *Kreis und Kugel*, Leipzig (1916), pp. 115-116.

[9] Cf., e. g., W. Blaschke, *loc. cit.*, p. 116.

and

(5) $\rho_1(\theta) \equiv r + h_n(\theta)$ if $\theta_{2n+1} > \theta > \theta_{2n+2}$; $(n = 0, 1, 2, \cdots)$

where $h_n(\theta) > 0$ is a function [10] (defined only on the interval $\theta_{2n+1} > \theta > \theta_{2n+2}$) for which all derivatives (and the function values) exist, and approach zero as $\theta \to \theta_{2n+1} - 0$ or $\theta \to \theta_{2n+2} + 0$, and such that the first n derivatives are less than $(\theta_{2n+2})^2$ in absolute value. Thus $\rho_1(\theta)$ is defined by (3), (4), (5) for the interval $0 < \theta \leqq \pi/2$. Let the definition of $\rho_1(\theta)$ be completed by the requirement

(6) $\rho_1(\theta)$ is periodic of period $\pi/2$.

Now $\rho_1(\theta)$ is differentiable infinitely often. This is obvious in the interval $0 < \theta \leqq \pi/2$. Thus, by (6), it is sufficient to show that all derivatives exist at the point $\theta = 0$. But, by (4) for $n = 0$, the left-hand derivatives at $\theta = \pi/2$ and hence, by (6), at $\theta = 0$ are all zero. On the other hand, by (3), (4), (5) and the definition of $h_n(\theta)$, the right-hand derivatives are all zero also. For suppose it has been proved that the k-th derivative at $\theta = 0$ exists and is zero, then the $(k+1)$-th difference quotients are bounded by $(\theta_{2n+2})^2/\theta_{2n+2} \to 0$.

Now, by (6), the closure conditions (1) are obviously satisfied so that there exists a convex curve C_1 of which $\rho_1(\theta)$ is the radius of curvature. The circle C_2 of radius r can be placed entirely within C_1 since $\rho_1(\theta) \geqq r$. Thus the vectorial sum $C_1(+)C_2$ does have an inner boundary curve C_I. But if the "mechanical" interpretation of vectorial addition mentioned above be remembered, it is clear that each of the intervals involved in (4) will be directions of distinct corners of this inner boundary C_I. This completes the proof of Theorem II.

It is noticed that the example was so constructed that the points of C_1 and C_2 which corresponded to a cluster point of corners of C_I were points where the radii of curvature of the two curves were equal. This fact could not be avoided. In fact, it is a direct consequence of Lemma I that if the two curves C_1 and C_2 have continuous radii of curvature $\rho_1(\theta)$ and $\rho_2(\theta)$ and if the inner boundary curve of $C_1(+)C_2$ exists and has infinitely many corners clustering in the direction θ_0 then $\rho_1(\theta_0) = \rho_2(\theta_0 + \pi)$. Thus

THEOREM III. *If C_1 and C_2 have continuous radii of curvature and if the vectorial sum $C_1(+)C_2$ has an inner boundary curve C_I then the corners (if any) of C_I are nowhere dense on the inner curve.*

THE JOHNS HOPKINS UNIVERSITY.

[10] Such a function may be taken in the form

$$h_n(\theta) = k_n \exp[(\theta_{2n+1} - \theta)(\theta_{2n+2} - \theta)]$$

where the constant $k_n > 0$ is chosen sufficiently small.

REAL CANONICAL BINARY TRILINEAR FORMS.*

By Rufus Oldenburger.

1. Introduction. In 1922, E. Schwartz [1] found all of the canonical binary trilinear forms for the class of all non-singular linear transformations in the complex field, and distinguished them by means of algebraic invariants. In 1932, the author found these canonical forms independently, and classified them more briefly according to arithmetic invariants.[2] In the present paper the author obtains all of the canonical binary trilinear forms for the class of all non-singular linear transformations in the field of reals, and a complete invariant system. The number of canonical forms is finite and is one more than the number of such forms for the complex field. The method of treatment depends on the use of arithmetic invariants.

Explicitly, the problem solved in this paper is the following: Given two sets of real constants $\bar{a}_{rst}, a_{rst}$, $r, s, t = 1, 2$, find the conditions on $\bar{a}_{rst}, a_{rst}$, such that there exist real solutions $p_r{}^\rho, q_s{}^\sigma, m_t{}^\tau$ of the equations

$$\bar{a}_{rst} = a_{\rho\sigma\tau} p_r{}^\rho q_s{}^\sigma m_t{}^\tau, \qquad (r, s, t, \rho, \sigma, \tau = 1, 2),$$

for which the determinants $|p_r{}^\rho|$, $|q_s{}^\sigma|$, $|m_t{}^\tau|$ are not zero.

2. Definitions. In another paper the author [3] defined and made a thorough study of ranks of n-way matrices and associated forms. A few of these definitions are given here. The *rank* r_i of a 3-way matrix $A = (a_{ijk})$, $i, j, k = 1, 2$, and its associated trilinear form $F = a_{ijk} x_i y_j z_k$; $i, j, k = 1, 2$; is the rank of the 2-way matrix

$$\begin{pmatrix} a_{111} & a_{112} & a_{121} & a_{122} \\ a_{211} & a_{212} & a_{221} & a_{222} \end{pmatrix}.$$

The ranks r_j, r_k are defined similarly. Assume that a_{ijk} $(i, j, k = 1, 2)$ are not

* Received June 16, 1936; revised November 11, 1936.

[1] E. Schwartz, "Über binäre trilineare Formen," *Mathematische Zeitschrift*, vol. 12 (1922), pp. 18-35.

[2] R. Oldenburger, "On canonical binary trilinear forms," *Bulletin of the American Mathematical Society*, vol. 38 (1932), pp. 385-387. In this paper a complete bibliography of earlier papers on binary trilinear forms is given.

[3] R. Oldenburger, "Composition and rank of n-way matrices and multilinear forms," *Annals of Mathematics*, vol. 35 (1934), pp. 622-657.

all zero. The 3-way *rank* $r[jk, i]$ of A and F is defined to be 2 or 1 according as the quantities.[4]

$$(1) \qquad \begin{vmatrix} a_{111} & a_{112} \\ a_{121} & a_{122} \end{vmatrix}, \qquad \begin{vmatrix} a_{211} & a_{212} \\ a_{221} & a_{222} \end{vmatrix}, \qquad \begin{vmatrix} a_{111} & a_{212} \\ a_{121} & a_{222} \end{vmatrix} + \begin{vmatrix} a_{211} & a_{112} \\ a_{221} & a_{122} \end{vmatrix}.$$

are not all zero or are all zero. The ranks $r[ij, k]$, $r[ik, j]$ are defined similarly. Evidently $r[jk, i] = r[kj, i]$. These ranks are *invariant* under non-singular linear transformations on F.

In this paper two binary trilinear forms $F = a_{ijk}x_iy_jz_k$, $G = b_{pqr}x'_py'_qz'_r$ and their associated matrices will be said to be *equivalent* if there exist transformations

$$x_i = a_{ip}x'_p, \qquad y_j = b_{jq}y'_q, \qquad z_k = c_{kr}z'_r,$$

where the square matrices (a_{ip}), (b_{jq}), (c_{kr}) of the second order are non-singular, and these transformations bring F into G. Similarly if F, G are bilinear forms.

3. The canonical forms for which $r_i = r_j = r_k = 2$. By a theorem of another paper[5] if one of the ranks $r[ij, k]$, $r[jk, i]$, $r[ik, j]$ of $\boldsymbol{A}$ is 1, at least one of the ranks r_i, r_j, r_k of $\boldsymbol{A}$ is 1. Hence

$$r[ij, k] = r[jk, i] = r[ik, j] = 2.$$

Let $\boldsymbol{A}_1 = (a_{1jk})$, $\boldsymbol{A}_2 = (a_{2jk})$, $j, k = 1, 2$. Since the coefficients of ρ^2, $\rho\sigma$, σ^2 in the determinant $|\rho\boldsymbol{A}_1 + \sigma\boldsymbol{A}_2|$ are the quantities (1), it follows that

$$|\rho\boldsymbol{A}_1 + \sigma\boldsymbol{A}_2| \not\equiv 0.$$

If $\boldsymbol{A}_2$ is non-singular, while $\boldsymbol{A}_1$ is singular make the transformation $x_1 = x'_2$, $x_2 = x'_1$ on the form F. If $\boldsymbol{A}_1, \boldsymbol{A}_2$ are both singular and the matrix $(\rho\boldsymbol{A}_1 + \sigma\boldsymbol{A}_2)$ is non-singular, then $\rho\sigma \neq 0$. Let the bilinear forms $a_{1jk}y_jz_k$, $a_{2jk}y_jz_k$, $j, k = 1, 2$, be denoted by F_1, F_2 respectively. Let

$$F' = a_{ijk}x'_iy'_jz'_k = x'_1F'_1 + x'_2F'_2$$

denote a form for which

$$F'_1 = \rho F_1 + \sigma F_2, \qquad F'_2 = F_2,$$

where ρ, σ are chosen so that $|\rho\boldsymbol{A}_1 + \sigma\boldsymbol{A}_2| \neq 0$. Equating F' to $F = x_1F_1 + x_2F_2$, we obtain

---

[4] These are 3-way determinants of the second order.

[5] R. Oldenburger, *Annals of Mathematics*, vol. 35 (1934), p. 649.

$$(2) \qquad \begin{aligned} x_1 &= \rho x'_1, \\ x_2 &= \sigma x'_1 + x'_2. \end{aligned}$$

The non-singular transformation (2), therefore, reduces F to a form F' for which (a'_{1jk}) is non-singular. Since in every case F is equivalent to a form $F' = x'_1 F'_1 + x'_2 F'_2$, where F'_1 is non-singular, it is no restriction to assume in what follows that F_1 is non-singular.

The pair of bilinear forms F_1, F_2 is now equivalent in the field of reals to the canonical pair [6]

$$(3) \qquad F_1 = y_1 z_1 + y_2 z_2, \qquad F_2 = y_1 z_2 + a y_2 z_1 + b y_2 z_2.$$

It is to be noted that any pair of binary bilinear forms is rationally equivalent, in the non-singular case, to the pair (3) or to the pair

$$(4) \qquad y_1 z_1 + y_2 z_2, \qquad a(y_1 z_1 + y_2 z_2).$$

Now F_1, F_2 are not equivalent to (4), since, then, the form $F = x_1 F_1 + x_2 F_2$ has $r_i = 1$.

In what follows in this section, we shall assume that $F = x_1 F_1 + x_2 F_2$, where F_1, F_2 are as given in (3). We shall consider three cases.

Case 1. $b^2 + 4a > 0$. In the field of reals, the determinant $D = | \rho A_1 + \sigma A_2 |$ factors into distinct linear factors $(\alpha\rho + \beta\sigma)$, $(\gamma\rho + \delta\sigma)$. Let

$$(5) \qquad \rho' = \alpha\rho + \beta\sigma, \qquad \sigma' = \gamma\rho + \delta\sigma.$$

Then $D = \rho'\sigma'$. By another paper of the author [7] the transformation (5) corresponds to a non-singular linear transformation on the x's of F, giving a new form $F' = a'_{ijk} x'_i y_j z_k$, for which

$$(6) \qquad | \rho' a'_{1jk} + \sigma' a'_{2jk} | = \rho'\sigma'.$$

Since the coefficients of ρ'^2 and σ'^2 vanish in (6), the 2-way matrices (a'_{1jk}), (a'_{2jk}) are singular. The form $F'_1 = a'_{1jk} y_j z_k$ is evidently equivalent to a form

$$F_1'' = y_1'' z_1''.$$

Simultaneously $F_2' = a'_{2jk} y_j z_k$ transforms into a form

$$F_2'' = e y_1'' z_1'' + f y_1'' z_2'' + g y_2'' z_1'' + h y_2'' z_2''.$$

[6] L. E. Dickson, *Modern Algebraic Theories*, pp. 89-97.

[7] R. Oldenburger, *Transactions of the American Mathematical Society*, vol. **39** (1936), pp. 432-433.

Hence F' is equivalent to

$$F'' = x_1''F_1'' + x_2''F_2''.$$

Since F_2'' is singular, we can write

$$F_2'' = (\alpha y_1'' + \beta y_2'')(\gamma z_1'' + \delta z_2'').$$

If $\beta, \delta \neq 0$, making the non-singular transformations

$$\begin{aligned} y_1''' &= y_1'', & y_2''' &= \alpha y_1'' + \beta y_2'', \\ z_1''' &= z_1'', & z_2''' &= \gamma z_1'' + \delta z_2'' \end{aligned}$$

on F'', we obtain the canonical form

$$R = x_1 y_1 z_1 + x_2 y_2 z_2,$$

where, for simplicity, the primes on the variables have been removed.

If $\beta = 0$, we can write

$$F'' = y_1''B,$$

where B is a bilinear form in the x's and z's; whence the rank r_j of F'' is 1. Similarly if $\delta = 0$

$$F'' = z_1''Q,$$

where Q is bilinear in the x's and y's and r_k of F'' is 1. In either case we obtain a contradiction of the assumption $r_j = r_k = 2$.

Case 2. $b^2 + 4a = 0$. In this case, $|\rho A_1 + \sigma A_2|$ is a perfect square. The form $F = x_1 F_1 + x_2 F_2$ defined by (3) is now

$$F = x_1(y_1 z_1 + y_2 z_2) + x_2(y_1 z_2 - g^2 y_2 z_1 + 2g y_2 z_2),$$

where $g = b/2$. Assume that $b \neq 0$. Making the non-singular transformations

$$\begin{aligned} x_1 &= g(x'_2 - x'_1), & x_2 &= x'_1, \\ y_1 &= y'_2, & y_2 &= -(y'_1 + y'_2)/g, \\ z_1 &= (z'_1 - z'_2)/g, & z_2 &= -z'_2 \end{aligned}$$

on F, we obtain

$$L = x_1 y_1 z_1 + x_2 y_1 z_2 + x_2 y_2 z_1,$$

where we have dropped the primes in L.

If $b = 0$, then $a = 0$. Interchanging x_1, x_2 and z_1, z_2 in F, we obtain L.

Case 3. $b^2 + 4a < 0$. In this case, $|\rho A_1 + \sigma A_2|$ does not factor in the field of reals. Let

$$M = x'_1 F'_1 + x'_2 F'_2,$$

where

(7) $$F'_1 = y'_1 z'_1 + y'_2 z'_2, \qquad F'_2 = y'_1 z'_2 - y'_2 z'_1.$$

We shall prove that $F = x_1 F_1 + x_2 F_2$, defined by (3), is equivalent to M. Apply to M the non-singular linear transformation

(8) $$x'_1 = \rho x_1 + \tau x_2, \qquad x'_2 = \sigma x_1 + \xi x_2.$$

This gives the new form

$$M' = x_1(\rho F'_1 + \sigma F'_2) + x_2(\tau F'_1 + \xi F'_2).$$

For F to be equivalent to M in the field of reals, it is evidently necessary and sufficient that there exist real quantities ρ, σ, τ, ξ, such that, if we write

$$\Delta = \begin{vmatrix} \rho & \sigma \\ \tau & \xi \end{vmatrix},$$

then

(9) $$\Delta \neq 0,$$

and there exist real non-singular transformations on the y's and z's so that M' becomes F. Then there must exist real values of ρ, σ, τ, ξ satisfying (9) such that the pair of bilinear forms $\rho F'_1 + \sigma F'_2$, $\tau F'_1 + \xi F'_2$ is equivalent under non-singular transformations on the y's and z's in the field of reals to F_1, F_2. This equivalence is satisfied *if and only if these pairs of forms have the same invariant factors.*[8] Let

$$\boldsymbol{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \boldsymbol{A} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \qquad \boldsymbol{B} = \begin{pmatrix} 0 & 1 \\ a & b \end{pmatrix}.$$

The characteristic matrix of F_1, F_2 is

$$(\lambda \boldsymbol{I} + \mu \boldsymbol{B}) = \begin{pmatrix} \lambda & \mu \\ a\mu & \lambda + b\mu \end{pmatrix},$$

which has the unique invariant factor

(10) $$|\lambda \boldsymbol{I} + \mu \boldsymbol{B}| = \lambda^2 + b\lambda\mu - a\mu^2.$$

The characteristic determinant of $\rho F'_1 + \sigma F'_2$, $\tau F'_1 + \xi F'_2$ is

(11) $$|\lambda(\rho \boldsymbol{I} + \sigma \boldsymbol{A}) + \mu(\tau \boldsymbol{I} + \xi \boldsymbol{A})| = \lambda^2 Q(\rho, \sigma) + 2\lambda\mu B(\rho, \sigma, \tau, \xi) + \mu^2 Q(\tau, \xi),$$

where $\quad Q(\rho, \sigma) = \rho^2 + \sigma^2, \qquad B(\rho, \sigma, \tau, \xi) = (\rho\tau + \sigma\xi).$

---

[8] L. E. Dickson, *Modern Algebraic Theories*, p. 115.

The invariant factor (10) is equal to the corresponding invariant factor of $\rho F'_1 + \sigma F'_2, \tau F'_1 + \xi F'_2$ if and only if the coefficients of (11) are proportional to those of (10). Then there must exist real values of ρ, σ, τ, ξ satisfying (9), and a real $k \neq 0$ such that

$$(12) \qquad Q(\rho, \sigma) = k, \qquad Q(\tau, \xi) = -ka, \qquad B(\rho, \sigma, \tau, \xi) = kb/2.$$

Since $b^2 + 4a < 0$, we have $a < 0$. It follows, since $Q(\rho, \sigma)$ is a positive definite quadratic form, that $Q(\rho, \sigma)$ represents $k, -ka$ where $k > 0$. If $\rho_1, \sigma_1, \tau_1, \xi_1$ are a set of real values satisfying (12) for a given real $k \neq 0$, the real quantities

$$\rho = \frac{\rho_1}{\sqrt{k}}, \qquad \sigma = \frac{\sigma_1}{\sqrt{k}}, \qquad \tau = \frac{\tau_1}{\sqrt{k}}, \qquad \xi = \frac{\xi_1}{\sqrt{k}}$$

satisfy (12), where in (12) we set $k = 1$. We therefore restrict our study of solutions of (12) to the case $k = 1$. Solving (12_1), (12_2) with $k = 1$ we obtain

$$(13) \qquad \rho = \pm\sqrt{1 - \sigma^2}, \qquad \tau = \pm\sqrt{-a - \xi^2}.$$

Substituting the solutions (13) in (12_3) with $k = 1$, we obtain

$$(14) \qquad \pm\sqrt{(1 - \sigma^2)(-a - \xi^2)} = (b - 2\sigma\xi)/2.$$

Set

$$(15) \qquad \xi = k_1\sqrt{-a}.$$

Since $a < 0$, ξ is real if k_1 is real. Assume henceforth that k_1 is real. Substituting (15) in (14), we obtain the following solution

$$(16) \qquad \sigma = \frac{-bk_1 \pm \sqrt{(b^2 + 4a)(k_1^2 - 1)}}{-2\sqrt{-a}},$$

and from (13) the solution

$$(17) \qquad \tau = \pm\sqrt{a(k_1^2 - 1)}.$$

Since $b^2 + 4a$, $a < 0$, the quantities σ, τ are real if and only if

$$(18) \qquad k_1^2 \leqq 1.$$

Substituting for ρ, ξ, τ from (13_1), (15), (17) in Δ as given above (9), we find that $\Delta \neq 0$ if

$$\sigma^2 \neq k_1^2.$$

Substituting (16) in the relation

$$\sigma = \pm k_1,$$

transposing terms, squaring, and simplifying, we obtain

$$(19) \qquad k_1^2(4\sqrt{-a} \pm b) = -\frac{(b^2+4a)}{2\sqrt{-a}}.$$

Since the right member of (19) is not zero, if there exist solutions of (19), the left member is also $\neq 0$ and, for a given value of the $\pm$ sign, (19) is of the form

$$\alpha k_1^2 = \beta, \qquad \alpha, \beta \neq 0,$$

which has at most two real solutions for k_1.

By (13_1), ρ is real if and only if $\sigma^2 \leqq 1$; whence, by (16), assuming that $k_1^2 \leqq 1$, so that σ is real,

$$(20) \qquad -1 \leqq \frac{bk_1 \pm \sqrt{(b^2+4a)(k_1^2-1)}}{2\sqrt{-a}} \leqq 1.$$

Taking the value of the $\pm$ sign in (20) to be $+$, the right inequality of (20) can be reduced to

$$(21) \qquad \sqrt{(b^2+4a)(k_1^2-1)} \leqq 2\sqrt{-a} - bk_1.$$

If $b > 0$, the right member of (21) is $\geqq 0$ for

$$(22) \qquad \frac{2\sqrt{-a}}{b} \geqq k_1,$$

and, if $b < 0$, that member is $\geqq 0$ for

$$(23) \qquad \frac{2\sqrt{-a}}{b} \leqq k_1.$$

If the right member of (21) is $\geqq 0$, and $k_1^2 \leqq 1$, we can square both sides of (21). Simplifying the resulting inequality, we obtain

$$(b - 2\sqrt{-a}\ k_1)^2 \geqq 0,$$

which is satisfied for every real value of k_1.

If $b = 0$, $\sigma = \pm\sqrt{1-k_1^2}$, and $\rho = \pm k_1$, whence ρ is real for every real value of k_1.

Evidently, there is an unlimited number of real values of k_1 satisfying (18), (22) or (23), and not satisfying (19). Also, for any solutions of σ, ξ from (15), (16), the $\pm$ signs in ρ, τ can always be chosen so that (14) is satisfied. We have now proved that in every case we can chose k_1 so that

ρ, σ, τ, ξ are real, $\Delta \neq 0$, and (12) is satisfied. *Hence F is equivalent to M for all a, b such that $b^2 + 4a < 0$.*

4. The canonical forms for which $r_i = 1$. Assume that $r_i = 1$, $r_j = r_k = 2$. We can reduce the form $F = a_{ijk}x_i y_j z_k$, $i, j, k = 1, 2$, at once to $x_1 B$ where B is bilinear in y and z and of rank 2. Reducing B to canonical form we obtain

$$H = x_1 y_1 z_1 + x_1 y_2 z_2.$$

Assume that $r_i = r_j = r_k = 1$. F can be reduced at once to

$$K = x_1 y_1 z_1.$$

No form with $r_i = r_j = 1$, $r_k = 2$ exists. We have therefore treated all cases.

5. Fundamental theorems of equivalence. We have proved

THEOREM 1. *Two binary trilinear forms $F = a_{ijk}x_i y_j z_k$ and $G = b_{ijk}x'_i y'_j z'_k$ are equivalent in the field of reals, if and only if they have the same ranks* r_i, r_j, and r_k, and, if $r_i = r_j = r_k = 2$, *the determinants* $| \rho a_{1jk} + \sigma a_{2jk} |$ *and* $| \rho b_{1jk} + \sigma b_{2jk} |$ *have both*

(a) *distinct real linear factors,*

or

(b) *coincident real linear factors,*

or

(c) *no real linear factors.*

THEOREM 2. *In the field of reals, a binary trilinear form $F = a_{ijk}x_i y_j z_k$ is equivalent to one of the following canonical forms:*

(a) $R = x_1 y_1 z_1 + x_2 y_2 z_2$, *if $r_i = r_j = r_k = 2$, and part* (a) *of Theorem* 1 *is satisfied;*

(b) $L = x_1 y_1 z_1 + x_2 y_1 z_2 + x_2 y_2 z_1$, *if $r_i = r_j = r_k = 2$, and part* (b) *of Theorem* 1 *is satisfied;*

(c) $M = x_1 y_1 z_1 + x_1 y_2 z_2 + x_2 y_1 z_2 - x_2 y_2 z_1$, *if $r_i = r_j = r_k = 2$, and part* (c) *of Theorem* 1 *is satisfied;*

(d) $H = x_1 y_1 z_1 + x_1 y_2 z_2$, *if $r_i = 1$, $r_j = r_k = 2$.*

(e) $K = x_1 y_1 z_1$, *if $r_i = r_j = r_k = 1$.*

6. Note concerning M. In the theory of forms, an arithmetic invariant called "factorization rank" plays an important rôle. The factorization ranks

of R, L, H, K have been studied elsewhere[9] by the author. The factorization rank of M is 3, since the matrix (m_{ijk}) of M can be written in the form

$$(m_{ijk}) = \left(\sum_{\alpha=1}^{3} a_{\alpha i} b_{\alpha j} c_{\alpha k}\right), \tag{24}$$

where

$$(a_{\alpha i}) = \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad (b_{\alpha j}) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}, \qquad (c_{\alpha k}) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ -1 & 1 \end{pmatrix},$$

and not in the form (24), where the range of α is 1, 2.

7. *Reductions.* The transformations *reducing any trilinear form to canonical form for the field of reals can be written down at once* from the theory of this paper and known theory of bilinear forms.

ARMOUR INSTITUTE OF TECHNOLOGY.

---

[9] R. Oldenburger, "On arithmetic invariants of binary cubic and binary trilinear forms," *Bulletin of the American Mathematical Society*, vol. 42 (1936), pp. 871-873.

A REMARK ON A THEOREM OF ARZELÀ.*

By PHILIP HARTMAN.

Let I denote a bounded interval, and $\{f_n(x)\}$ a sequence of functions defined on I such that (i) $\{f_n(x)\}$ is uniformly bounded on I and (ii) every $f_n(x)$ is continuous on I. Condition (i) implies that for every enumerable subset C of I there exists a subsequence of $\{f_n(x)\}$ which is convergent at every point of C. Since C may be chosen dense on I, it follows from a standard theorem of Arzelà that if (i) is satisfied and (ii) is replaced by the more stringent condition that $\{f_n(x)\}$ be equicontinuous on I, then $\{f_n(x)\}$ contains a subsequence which is uniformly convergent on I. The question now arises whether or not (i) and (ii) alone imply the existence of a subsequence which is convergent on I. This question will be answered in the negative by proving a sharper statement to the effect that *there exist sequences* $\{f_n(x)\}$ *which satisfy* (i) *and* (ii) *but are such that every subsequence of* $\{f_n(x)\}$ *is divergent almost everywhere.* For instance, *every* subsequence of the sequence

$$\sin x, \sin 2x, \cdots, \sin nx, \cdots$$

will be shown to be divergent almost everywhere.

Let $\{k_n\}$ be any increasing sequence of positive integers. A theorem of Hardy and Littlewood (*Acta Mathematica*, vol. 37 (1914), p. 181) states that there exists a set $S = S(\{k_n\})$ of measure 1 in $[0, 1]$ such that if θ is a point of S then the sequence of numbers $\{(k_n\theta)\}$, where $(k_n\theta)$ denotes the fractional part of $k_n\theta$, is dense in $[0, 1]$. It follows that the sequence $\{\sin 2\pi k_n x\}$ is divergent at each point of S; for if θ is a point of S, the sequence of numbers $\{\sin 2\pi k_n \theta\}$ is dense in $[-1, 1]$.

Obviously, the above remarks also apply if $\sin x$ is replaced by any other non-constant, continuous, periodic function $f(x)$ and $f_n(x)$ is defined to be $f(nx)$.

THE JOHNS HOPKINS UNIVERSITY.

* Received January 6, 1937.

American Journal of Mathematics
CALCUTTA CENTRAL LIBRARY

ON THE REPRESENTATIONS OF THE SYMMETRIC GROUP.*

By F. D. MURNAGHAN.

Introduction. The representation theory of the symmetric group (group of $n!$ permutations of n letters) was initiated by Frobenius some forty years ago and was developed in the, now classical, papers of Schur and Young. More recently Littlewood and Richardson (**13**) have discussed in detail the problem of the construction of the character table and have used a recurrence formula (passing from the symmetric group on $(n-1)$ letters to the symmetric group on n letters) due to Schur in order to determine the characters of those classes of permutations which contain *at least one* unary cycle (= fixed letter). We show in the present paper that this recurrence formula of Schur is but a special case of a general recurrence formula by means of which the characters of a class containing *at least one cycle* on p letters $(1 \leq p < n)$ may be determined from the characters of the symmetric group on $n-p$ letters. As the characters of the class containing just one cycle (on n letters) are trivially evident (as was pointed out by Frobenius) the construction of the character tables for the various symmetric groups $(n=1, 2, 3, \cdots)$ is a routine matter demanding only paper and ink; and the easiest characters to calculate are those of classes containing cycles on the greatest number of letters.

The representation theory of the symmetric group is of importance in nuclear physics and in this connection the following two questions are of particular significance.

1. If we imagine the n letters, whose permutations constitute the symmetric group, to be divided up in compartments or boxes containing, respectively, $\lambda_1, \lambda_2, \cdots, \lambda_k$ letters (so that $\lambda_1 + \lambda_2 + \cdots + \lambda_k = n$) we obtain a

* Received May 31, 1937. This paper is an elaboration of an address delivered April 12, 1937, at the Institute for Advanced Study during the author's stay there as guest member 1936-37.

subgroup of the symmetric group by considering those permutations which do not send any letter out of its box. The cosets (right or left) of this subgroup furnish a representation (whose elements are permutation matrices) of the symmetric group; this representation is, in general, reducible and it is important to determine its analysis into irreducible components. The solution of this question is quite simple and is well known when $k = 2$, i. e. when there are but two boxes. We give the solution in the general case.

2. The "direct product" of an irreducible representation of the symmetric group on n letters by an irreducible representation of the symmetric group on m letters furnishes a representation, in general reducible, of the symmetric group on $n + m$ letters and it is important to determine the analysis of this reducible representation into its irreducible components. We show how to do this, without having to use the character tables, and record the results for all values of n and m for which $n + m \leq 9$.

In the hope of making the theory of the representations of the symmetric group more accessible to workers in nuclear physics we have made the following account somewhat self-contained. The original papers of Frobenius, and particularly those of Schur, arouse in a persevering reader an emotion akin to that inspired by one of the great symphonies; but they are by no means easy reading and we hope that a somewhat elementary orchestration may acquaint a larger audience with the work of the masters. It is a pleasure to here record our obligation, amongst others, to Professor Wedderburn for a pregnant remark which materially aided and simplified our treatment of the problem 1 of the preceding paragraph.

1. The characteristics of a finite group. Let G be a finite group of order N; its elements will fall into r classes (of conjugate elements) $C_1, \cdots, C_r$ such that if $g_r \subset C_r$ each element of C_r is of the form $g g_r g^{-1}$, $g \subset G$. We denote by N_p the number of elements in the p-th class C_p so that, C_1 being the class consisting of the identity element g_1, $N_1 = 1$ and $N_1 + N_2 + \cdots + N_r = N$. By a representation of G is meant a linear group homomorphic to G and it is well known that G possesses exactly r non-equivalent irreducible representations $\Gamma_1, \cdots, \Gamma_r$. These are distinguished from one another by their characters and we denote by $\chi_p{}^q$ the character of Γ_p associated with the class C_q; i. e. $\chi_p{}^q = \chi_p(g_q)$ where $g_q \subset C_q$. These characters satisfy certain fundamental orthogonality relations which are most conveniently stated as follows. If $a(g)$, $b(g)$ are any two complex valued functions defined over G we denote by $(\boldsymbol{a} \cdot \boldsymbol{b})$ the average of the product $a(g)\bar{b}(g)$ over G (the superposed bar denoting, as usual, the complex conjugate):

$$(\boldsymbol{a}\cdot\boldsymbol{b})=\frac{1}{N}\sum_{g}a(g)\bar{b}(g)$$

We shall be interested only in the case where $a(g)$ and $b(g)$ are class functions: $a(g_q)=a^q$; $b(g_q)=b^q$; $g_q\subset C_q$ and then $(\boldsymbol{a}\cdot\boldsymbol{b})=\frac{1}{N}\sum_{q}N_q a^q\bar{b}^q$. Then the orthogonality relations referred to are

$$(1)\qquad (\boldsymbol{\chi}_p\cdot\boldsymbol{\chi}_q)=0;\ p\neq q;\ (\boldsymbol{\chi}_p\cdot\boldsymbol{\chi}_p)=1;\qquad (p=1,2,\cdots,r).$$

These imply that any class function $\boldsymbol{a}$ is a linear combination of the r functions $\boldsymbol{\chi}_p$: $\boldsymbol{a}=c^a\boldsymbol{\chi}_a$ (usual summation convention) where $c^p=(\boldsymbol{a}\cdot\boldsymbol{\chi}_p)$. Denoting by $\boldsymbol{s}$ a class function whose r components s^q, $(q=1,\cdots,r)$, are indeterminates the expression $\phi(\boldsymbol{s})=(\boldsymbol{s}\cdot\boldsymbol{\chi}_p)=\frac{1}{N}\sum_{q}N_q\bar{\chi}_p{}^q s^q$ is called the *characteristic* of the irreducible representation Γ_p and the characters $\chi_p{}^q$ are called the components of this characteristic. Any representation of G is of the form $c^a\Gamma_a$ where the coefficients c^p are integers (positive or zero) and the characters of this representation are $c^a\chi_a{}^q$; the corresponding expression $c^a\phi_a(\boldsymbol{s})$ being termed the characteristic of the given representation (with components $c^a\chi_a{}^q$). When all the coefficients c^p vanish save one which is unity, so that the representation is irreducible, the characteristic is termed *simple*; otherwise it is called *compound* so that the characteristic of a reducible representation is compound. It is occasionally convenient to allow the coefficients c^p in the expression $c^a\phi_a(\boldsymbol{s})$ to take negative integral as well as positive integral or zero values and then $c^a\phi_a(\boldsymbol{s})$ is termed a generalized characteristic, it being clearly understood that if any of the coefficients c^p are negative the components $c^a\chi_a{}^q$ of the generalised characteristic are not the characters of any representation.

If $(\boldsymbol{s}\cdot c^a\boldsymbol{\chi}_a)$ is a generalised characteristic with components $c^a\chi_a{}^q=a^q$ we see at once from the orthogonality relations (1) that $(c^a\chi_a\cdot c^a\chi_a)=\sum_{a=1}^{r}(c^a)^2$ so that $(\boldsymbol{a}\cdot\boldsymbol{a})=1$ implies all the coefficients c^p zero save one which $=\pm 1$. The generalised characteristic will, therefore, be simple if $(\boldsymbol{a}\cdot\boldsymbol{a})=1$ and if a^1, the coefficient of $s\frac{1}{N}$, >0 (for the coefficient of s^1 in a simple characteristic yields, on multiplication by N, the dimension of the corresponding irreducible representation of G, and is, accordingly, positive). Amongst the irreducible representations $(\Gamma_1,\cdots,\Gamma_r)$ of *any* finite group occurs the identity representation Γ_1 (in which to each element $g\subset G$ there corresponds the one

dimensional unit matrix) and the associated simple characteristic is called the *principal* characteristic of G; its explicit expression is $\phi_1(\boldsymbol{s}) = \frac{1}{N}\sum_{q=1}^{r} N_q s^q$ so that the coefficient of s^q in $\phi_1(\boldsymbol{s})$ yields, on multiplication by N, the order of G, the number of elements in the class C_q. Finally the orthogonality relations (1) express the fact that the numbers $u_p{}^q = \sqrt{N_q/N}\cdot\chi_p{}^q$ are the elements of an $r \times r$ unitary matrix; so that

$$(2) \qquad \sum_{a=1}^{r} \chi_a{}^q \bar{\chi}_a{}^p = 0;\ q \neq p; \qquad \sum_{a=1}^{r} \chi_a{}^q \bar{\chi}_a{}^q = N/N_q.$$

Hence the equations $\phi_p(\boldsymbol{s}) = \frac{1}{N}\sum_q N_q \bar{\chi}_p{}^q s^q$ may be solved for the indeterminates s^q the solution being

$$(3) \qquad s^j = \sum_a \chi_a{}^j \phi_a(\boldsymbol{s}).$$

Before passing to our subject proper, the symmetric group, it is necessary to say a few words concerning a basic theorem of Frobenius which enables us to derive from a characteristic of a subgroup H of G a characteristic of G itself. Let H be of order M and denote by h a typical element of H; the class of H to which h belongs is a subset, proper or not, of the class of G to which h belongs. But a class C_j of G may contain several classes of H or none at all; we say that H refines the classes of G. If Γ is any representation of G it induces a representation Γ^- of H where Γ^- consists of those linear operators of Γ which remain after the operators which correspond to elements of G which are not in H are rejected. If, in particular, Γ is an irreducible representation of G Γ^- will be, in general, a *reducible* representation of H(since it may be possible to find a proper, non-trivial subspace of the carrier space of Γ which is invariant under all the operators of Γ^- although the irreducibility of Γ guarantees that no such subspace exists which is invariant under all the operators of Γ). If we have any class function $a(g)$ defined over G it induces by the process of projection: $a^*(h) = a(h)$ a class function $a^*(h)$ defined over H. Since $a(g) = \sum_{a=1}^{r} (\boldsymbol{a}\cdot\boldsymbol{\chi}_a)\chi_a(g)$ we have $a^*(h) = \sum_{a=1}^{r} (\boldsymbol{a}\cdot\boldsymbol{\chi}_a)\chi_a(h)$ or, equivalently, $\boldsymbol{a}^* = \sum_{a=1}^{r} (\boldsymbol{a}\cdot\boldsymbol{\chi}_a)\boldsymbol{\chi}^*_a$. In particular when $\boldsymbol{a}$ is the indeterminate $\boldsymbol{s}$ whose components s^q appeared in the definition of the simple characteristics of G we have $\boldsymbol{s}^* = \sum_{a=1}^{r} \phi_a(\boldsymbol{s})\boldsymbol{\chi}^*_a$ where the numbers $\chi^*{}_p{}^j$ are the characters of a representation (in general reducible) of H the index j running over the classes

of H. If these number t and if the characters of the irreducible representations of H be denoted by $\xi_k{}^j\ (k, j = 1, \cdots, t)$, we may write $\chi^*{}_p = c_p{}^a \xi_a$ where the coefficients $c_p{}^j$ are positive integers. The expression $(s^* \cdot d^\beta \xi_\beta) = d^\beta (s^* \cdot \xi_\beta)$ where the coefficients d^j are integers, positive, negative or zero, is a generalised characteristic of H, it being clearly understood that the indeterminates $s^*(h)$ are conditioned by the fact that they are the same for all elements of H which lie in the same class of G (and not merely the same for all elements of H which lie in the same class of H). On substituting for s^* its expression given above this generalised characteristic of H appears in the form

$$\sum_{a=1}^{r} \phi_a(s) d^\beta (\chi^*{}_a \cdot \xi_\beta) = \sum_{a=1}^{r} \phi_a(s) \sum_{\beta=1}^{t} (c_a{}^\beta d^\beta)$$

owing to the orthogonality relations amongst the characters of the irreducible representations of H. Since $\sum_{\beta=1}^{t} (c_a{}^\beta d^\beta)$ is an integer, positive, negative, or zero, it follows that *any* generalised characteristic of H furnishes by the procedure outlined a generalised characteristic of G (of which the components corresponding to classes of G which contain no elements of H are zero). As a trivial instance of this theorem let G be the symmetric group on 2 letters and H the identity element. The principal characteristic of H is s^1 and this being a generalised characteristic of G its components are $(2, 0)$ since $s^1 = \frac{1}{2}(2s^1 + 0 \cdot s^2)$. The generalised characteristic of G obtained in this way contains the principal characteristic $\phi_1(s) = \frac{1}{2}(s^1 + s^2)$ once since $c^1 = \frac{1}{2}(2.1 + 0.1) = 1$ and the remaining characteristic is $\frac{1}{2}(s^1 - s^2)$. This characteristic is simple since $\frac{1}{2}\{(1)^2 + (-1)^2\} = 1$ and, in addition, the coefficient of s^1 is positive. Thus the two simple characteristics of the symmetric group on two letters are $\phi_1(s) = \frac{1}{2}(s^1 + s^2)$ and $\phi_2(s) = \frac{1}{2}(s^1 - s^2)$ the corresponding characters being $(1, 1)$ and $(1, -1)$ respectively. As a less trivial example let G be the symmetric group on 3 letters; $N = 6$, $N_1 = 1$; $N_2 = 3$, $N_3 = 2$ and let H be the symmetric group on two of the three letters. The principal characteristic of H is $\frac{1}{2}(s^1 + s^2)$ and writing this in the form $\frac{1}{6}(3s^1 + 3s^2 + 0 \cdot s^3)$ its components (when viewed as a generalised characteristic of G) are $(3, 1, 0)$. It contains the principal characteristic $\phi_1(s) = \frac{1}{6}(s^1 + 3s^2 + 2s^3)$ of G once since $c^1 = \frac{1}{6}(1.3.1 + 3.1.1 + 2.0.1) = 1$ and the remaining characteristic is $\frac{1}{6}(2s^1 + 0.s^2 - 2.s^3)$ with components $(2, 0, -1)$. This characteristic is simple since $\frac{1}{6}(2^2 + 0^2 + 2.(-1)^2) = 1$ and the coefficient of s^1 is positive. We have, then, the two simple characteristics $\phi_1(s) = \frac{1}{6}(s^1 + 3s^2 + 2s^3)$; $\phi_2(s) = \frac{1}{6}(2s^1 + 0.s^2 - 2s^3)$ of the symmetric group on 3 letters. Starting

with the characteristic $\frac{1}{2}(s^1 - s^2)$ of H we find $c^1 = 0$, $c^2 = 1$ the remaining characteristic $\phi_3(s) = \frac{1}{6}(s^1 - 3s^2 + 2s^3)$ being simple. $\phi_1(s)$, $\phi_2(s)$, $\phi_3(s)$ are the three simple characteristics of the symmetric group on 3 letters the corresponding characters being $(1, 1, 1)$, $(2, 0, -1)$, and $(1, -1, 1)$ respectively.

We obtain a representation, in general reducible, of G in the following manner. Any subgroup H, of order M, of G has $d = N/M$ right cosets $H_1 = H$; $H_2 = Hg_2, \cdots, H_d = Hg_d$ and if we define $(H'_1, \cdots, H'_d)$ by the equation $H'_p = H_p g$ where g is an arbitrary element of G the symbols H' constitute a permutation of the symbols H; i. e. $\boldsymbol{H'} = P(g)\boldsymbol{H}$ where $P(g)$ is a permutation matrix. The matrices $P(g)$ furnish a representation of G; if g_q is any member of the class C_q of G $(q = 1, \cdots, r)$, the character χ^q associated with the class C_q in this representation is the number of ones in the diagonal of the permutation matrix $P(g_q)$ i. e. the number of elements g_p for which $H_p g_q = H_p$. In other words χ^q is the number of elements g_p for which $g_p g_q g_p^{-1} \subset H$, or, since this number is the same for each $g_q \subset C_q$, $\chi^q =$ (number of times $g_p C_q g_p^{-1}$ lies in H) $\div N_q$. But $g_p C_q g_p^{-1} = C_q$ so that as p runs from 1 to d we obtain $C_q : d = N/M$ times. Hence $\chi^q = N/M$ times the number of elements of $H \subset C_q \div N_q$. This suffices to show that the generalised characteristic of G obtained from the *principal* characteristic of H by the method of the previous paragraph is really a compound (or simple) characteristic; in fact the characteristic of that representation (by permutation matrices) which is furnished by the cosets of H in G. For the principal characteristic of H is $1/M \sum\limits_h s(h)$; written as a generalised characteristic of G it appears as $1/N \sum\limits_h Ns(h)/M$ so that its components $c^a \chi_a{}^q$ are N/M times the number of elements of H in $C_q \div N_q$. These being precisely the characters of the representation referred to, the theorem stated follows since two representations with the same characters or characteristic, are equivalent.

2. The principal and alternating characteristics of the symmetric group. Construction of reducible representations. Each permutation of the symmetric group on n letters may be written in a unique manner as a product of cycles, no letter appearing in more than one cycle. Two permutations with the same cycle structure, i. e. containing the same number α_1 of cycles on one letter (= unary cycles), the same number, α_2 of cycles on two letters (= binary cycles), the same number, α_3 of ternary cycles and so on, belong to the same class. For example if $n = 5$ and $P = (12)(345)$, $Q = (23)(154)$ the permutation $T = \begin{pmatrix} 12345 \\ 23154 \end{pmatrix} = (123)(45)$ transforms P into $Q : Q = TPT^{-1}$. We

refer to the class with the cycle structure $(\alpha) = (\alpha_1, \cdots, \alpha_n)$ as the class (α) and observe that $\alpha_1 + 2\alpha_2 + \cdots + n\alpha_n = n$ so that the number of classes, and hence of irreducible representations, is the number of solutions of this equation in non-negative integers. Writing

$$\alpha_1 + \alpha_2 + \cdots + \alpha_n = \lambda_1;\ \alpha_2 + \cdots + \alpha_n = \lambda_2;\ \cdots \alpha_n = \lambda_n$$

it is clear that

$$\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_n;\quad \lambda_1 + \lambda_2 + \cdots + \lambda_n = n.$$

If k is such that $\lambda_k > 0$, $\lambda_{k+1} = \lambda_{k+2} = \cdots = \lambda_n = 0$ we have

$$\lambda_1 + \lambda_2 + \cdots + \lambda_k = n;\ \lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_k > 0$$

so that the number of classes, or of irreducible representations, is the same as the number of partitions of n into sums of positive integers. We shall indicate a partition by the symbol $(\lambda_1, \lambda_2, \cdots, \lambda_k)$ where the parts $\lambda_1, \cdots, \lambda_k$ are written in non-increasing order and shall use an obvious exponent notation for the sake of brevity when two or more parts are equal. E. g. $(3, 2^2, 1^3)$ denotes the partition $(3, 2, 2, 1, 1, 1)$ of 10: $3 + 2 + 2 + 1 + 1 + 1 = 10$. Each partition is conveniently indicated by a diagram of horizontal rows of dots all beginning on the same vertical line; thus $(3, 2^2, 1^3)$ is indicated by the diagram

$$\begin{array}{lll} \cdot & \cdot & \cdot \\ \cdot & \cdot & \\ \cdot & \cdot & \\ \cdot & & \\ \cdot & & \\ \cdot & & \end{array}$$

By a simple interchange of the rows and columns of a diagram we obtain a second diagram (termed the *associate* of the original diagram) and hence a second partition of n (termed the associate of the original partition). E. g., the associate of $(3, 2^2, 1^3)$ is $(6, 3, 1)$. When the associate is identical with the original the diagram (and partition) are termed *self-associated.* E. g., $(3, 2, 1)$ is a self-associated partition of 6. We shall see how to attach to each diagram, or partition of n, a uniquely determined irreducible representation of the symmetric group and then the representations attached to associated

diagrams, or partitions of n, are termed associated; a representation attached to a self-associated diagram or partition being termed self-associated. We shall denote the partition associated with $(\lambda) = (\lambda_1, \lambda_2, \cdots, \lambda_k)$ by $(\mu) = (\mu_1, \cdots, \mu_j)$ and it is an immediate consequence of the definition that $\mu_1 = k$, $j = \lambda_1$; it is also clear that there are α_1 ones, α_2 twos, α_3 threes etc. in the partition (μ) where $\alpha_1 = (\lambda_1 - \lambda_2)$; $\alpha_2 = (\lambda_2 - \lambda_3), \cdots, \alpha_n = \lambda_n$ or, equivalently, $\alpha_1 + \alpha_2 + \cdots + \alpha_n = \lambda_1$; $\alpha_2 + \cdots + \alpha_n = \lambda_2$; $\cdots \alpha_n = \lambda_n$.

The number $N_{(\alpha)}$ of elements in the class (α) is readily found. If any such permutation is written down with the α_1 unary cycles appearing first, the α_2 binary cycles next, and so on, we obtain by mere permutation of the letters $n!$ permutations all in the class (α). But there are repetitions due to the fact that each cycle may begin, without changing it, with any one of its letters and to the fact that the α_r r-cycles may be permuted amongst themselves without affecting the permutation. Hence

$$N_{(\alpha)} = \frac{n!}{\prod_{p=1}^{p=n} p^{\alpha_p} \cdot \alpha_p!} = \frac{n!}{1^{\alpha_1} \cdot \alpha_1!\, 2^{\alpha_2} \cdot \alpha_2! \cdots n^{\alpha_n} \cdot \alpha_n!}.$$

If $(s_1, \cdots, s_n)$ are n indeterminates the expressions $s^{(\alpha)} = s_1^{\alpha_1} s_2^{\alpha_2} \cdots s_n^{\alpha_n}$ are class functions defined over the symmetric group and we use them in the definition of the simple characteristics of the symmetric group:

$$\phi_{(\lambda)}(s) = (s \cdot \chi_{(\lambda)}) = \frac{1}{n!} \sum_{(\alpha)} N_{(\alpha)} \bar{\chi}_{(\lambda)}{}^{(\alpha)} s^{(\alpha)}.$$

We shall denote the principal characteristic (i. e. the simple characteristic corresponding to the identity representation) by $q_n(s)$ so that

$$q_n(s) = \sum_{(\alpha)} \frac{1}{\alpha_1!\, \alpha_2! \cdots \alpha_n!} \left(\frac{s_1}{1}\right)^{\alpha_1} \left(\frac{s_2}{2}\right)^{\alpha_2} \cdots \left(\frac{s_n}{n}\right)^{\alpha_n}. \tag{4}$$

These polynomials $(n = 1, 2, \cdots)$ in the indeterminates $(s_1, s_2, \cdots)$ are the bricks with which will be built the characters of the irreducible representations of the symmetric group and we write out explicitly the first seven of them:

$$q_1(s) = s_1; \qquad q_2(s) = \tfrac{1}{2}(s_1^2 + s_2); \qquad q_3(s) = \frac{1}{3!}(s_1^3 + 3s_1s_2 + 2s_3)$$

$$q_4(s) = \frac{1}{4!}\{s_1^4 + 6s_1^2s_2 + 8s_1s_3 + 3s_2^2 + 6s_4\}$$

$$q_5(s) = \frac{1}{5!}\{s_1^5 + 10s_1^3s_2 + 20s_1^2s_3 + 15s_1s_2^2 + 30s_1s_4 + 20s_2s_3 + 24s_5\}$$

$$q_6(s) = \frac{1}{6!}\{s_1^6 + 15s_1^4s_2 + 40s_1^3s_3 + 45s_1^2s_2^2 + 90s_1^2s_4 + 120s_1s_2s_3 + 144s_1s_5 + 15s_2^3 + 90s_2s_4 + 40s_3^2 + 120s_6\}$$

$$q_7(s) = \frac{1}{7!}\{s_1^7 + 21s_1^5s_2 + 70s_1^4s_3 + 105s_1^3s_2^2 + 210s_1^3s_4 + 420s_1^2s_2s_3 + 504s_1^2s_5 + 105s_1s_2^3 + 630s_1s_2s_4 + 280s_1s_3^2 + 840s_1s_6 + 210s_2^2s_3 + 504s_2s_5 + 420s_3s_4 + 720s_7\}.$$

The terms are arranged so that $s_1^{m_1}s_2^{m_2}\cdots$ comes before $s_1^{n_1}s_2^{n_2}\cdots$ if the first non-vanishing number of the set $m_1 - n_1$, $m_2 - n_2, \cdots$ is positive. The polynomials $q_n(s)$ furnish at a glance the structure of the corresponding symmetric group. Thus from $q_6(s)$ we see that the $6! = 720$ permutations of the symmetric group on 6 letters divide into 11 classes there being 45 elements, for example, in the class $\alpha = (2, 2, 0, 0, 0, 0)$. It is clear from the defining formula (4) that the polynomials $q_n(s)$ satisfy the interconnecting relations

$$\frac{\partial q_n}{\partial s_1} = q_{n-1}; \quad \frac{\partial q_n}{\partial s_2} = \tfrac{1}{2}q_{n-2}; \quad \frac{\partial q_n}{\partial s_3} = \tfrac{1}{3}q_{n-3} \text{ and, generally,}$$

$$(5) \qquad \frac{\partial q_n}{\partial s_p} = \frac{1}{p}\, q_{n-p}, \qquad (p = 1, 2, \cdots, n)$$

where, to secure the universal validity of these formulae, we define $q_0(s) = 1$; $q_{-1}(s) = q_{-2}(s) = \cdots = 0$. We shall see that the relations (5) have the following significance; they enable us to construct, in a very simple manner, the characters of any class of the symmetric group on n letters which contains at least one cycle on p letters, $p = 1, \cdots, n - 1$, from the, supposed known, characters of the symmetric group on $(n - p)$ letters.

Since any cycle on p letters may be written as the product of $p - 1$ binary cycles ($=$ transpositions): E. g., $(1234) = (12)(13)(14)$, the order of the factors being from left to right: every cycle on an even number of letters is an odd permutation and every cycle on an odd number of letters is an even one. Hence all permutations in a given class (α) are either even or odd and we may speak of even or odd classes; a class (α) being even when $\alpha_2 + \alpha_4 + \alpha_6 + \cdots$ is even and odd when it is odd. Now the symmetric group on n letters possesses, in addition to the identity representation, a second one-dimensional representation; namely that one which attaches to each even permutation the number 1 and to each odd permutation the number -1 (this being merely a sophisticated way of saying that the product of two even or odd permutations is even whilst the product of an even by an odd permutation is odd). This representa-

tion is known as the alternating representation and the corresponding simple characteristic is known as the alternating characteristic; we shall denote it by $\pi_n(s)$ so that

$$\pi_n(s) = \sum_{(c)} \frac{(-1)^{\alpha_2+\alpha_4+\cdots}}{\alpha_1!\,\alpha_2!\cdots\alpha_n!}\left(\frac{s_1}{1}\right)^{\alpha_1}\left(\frac{s_2}{2}\right)^{\alpha_2}\cdots\left(\frac{s_n}{n}\right)^{\alpha_n}$$

implying

$$(6) \qquad \pi_n(s_1, s_2, \cdots, s_n) = q_n(s_1, -s_2, s_3, -s_4, \cdots).$$

Before passing to the question of associating with each partition (λ) of n a (reducible) representation of the symmetric group on n letters we find it convenient to remark that the characters $\chi_{(\lambda)}{}^q$ of the irreducible representations are all real (it appears in the sequel that they are all integers, positive, negative or zero, but this fact does not lie on the surface as does the fact of their reality). Indeed since the reciprocal of a cycle is the same cycle written in the reversed sense: E. g., $(1234)^{-1} = (1432)$: each class (α) contains the reciprocal of each of its permutations so that the character of any element is the same as the character of its reciprocal. But every representation of any *finite* group is equivalent to a representation by means of unitary matrices and, the reciprocal of a unitary matrix being its transposed conjugate, its trace is the conjugate complex of the trace of the original matrix. Hence the character $\chi(g^{-1})$ (for *any* representation of *any* finite group) is furnished by the relation $\chi(g^{-1}) = \bar{\chi}(g)$. For the *symmetric* group we have, in addition, the relation $\chi(g^{-1}) = \chi(g)$ and the two relations together imply $\bar{\chi}(g) = \chi(g)$ i. e. the reality of the characters of *any* representation of the symmetric group. We may, therefore, drop the conjugate complex sign in the explicit expressions for the simple characteristics and write

$$(7) \qquad \phi_{(\lambda)}(s) = \sum_{(\alpha)} \frac{\chi_{(\lambda)}{}^{(\alpha)}}{\alpha_1!\,\alpha_2!\cdots\alpha_n!}\left(\frac{s_1}{1}\right)^{\alpha_1}\cdots\left(\frac{s_n}{n}\right)^{\alpha_n}.$$

In order to associate with each partition (λ) of n a (reducible) representation of the symmetric group on n letters, we have merely to imagine the n letters placed in compartments or boxes containing, respectively, $\lambda_1, \lambda_2, \cdots, \lambda_k$ letters and then to consider the subgroup H of the symmetric group G which consists of those permutations which do not send any letter out of its box. This subgroup is of order $M = \lambda_1!\,\lambda_2!\cdots\lambda_k!$ and a typical permutation of it is of the form $P = P_1P_2\cdots P_k$ where P_j denotes a permutation on the letters of the j-th box $(j = 1, 2, \cdots, k)$. Since the various P_j operate on distinct letters the order in which the factors P_j are written is indifferent and we

agree to write them in the natural order $P_1P_2\cdots P_k$. If $(\alpha^j)=(\alpha_1{}^j,\cdots,\alpha_n{}^j)$ denotes the cycle structure of P_j then the cycle structure of P is furnished by the formula

$$\alpha_p=\sum_{j=1}^{k}\alpha_p{}^j; \qquad (p=1,2,\cdots,n).$$

Hence

$$s^{(\alpha)}=s_1{}^{\alpha_1}s_2{}^{\alpha_2}\cdots s_n{}^{\alpha_n}=\prod_{j=1}^{k}s^{(\alpha^j)}$$

and so the principal characteristic of H is

$$\frac{1}{\lambda_1!\,\lambda_2!\cdots\lambda_k!}\prod_{j=1}^{k}\sum_{j}s^{(\alpha_j)}$$

where $\sum_j$ denotes summation over the $\lambda_j!$ permutations on the letters in the j-th box. On writing this principal characteristic of H in the form

$$\prod_{j=1}^{k}\left(\frac{1}{\lambda_j!}\sum_{j}s^{(\alpha^j)}\right)$$

it appears as $q_{\lambda_1}(s)q_{\lambda_2}(s)\cdots q_{\lambda_k}(s)$. This product is, accordingly, a compound characteristic of the symmetric group G on n letters; namely, the characteristic of the representation of G furnished by the permutations of the cosets of H in G. Since the representation is by means of permutation matrices its characters are integers positive or zero and so the components of the characteristic (compound) $q_{\lambda_1}(s)\cdots q_{\lambda_k}(s)$ of G are integers positive or zero. We shall denote by $\Delta(\lambda_1,\lambda_2,\cdots,\lambda_k)$ the representation (reducible) of the symmetric group on n letters whose characteristic is $q_{\lambda_1}(s)\cdots q_{\lambda_k}(s)$.

$\Delta(\lambda_1,\cdots,\lambda_k)$ is sometimes referred to as a tensor representation for a reason that will be clear from the following examples:

1. If $k=2$, so that n is partitioned into two parts, $\Delta(\lambda_1,\lambda_2)$ is of dimension $\binom{n}{\lambda_2}=\frac{n!}{\lambda_1!\lambda_2!}$ and the cosets of H permute like the products of n letters $(x_1,\cdots,x_n)$ λ_2 at a time. These products may be regarded as a basis in a carrier "tensor" space of $\binom{n}{\lambda_2}$ dimensions in which $\Delta(\lambda_1,\lambda_2)$ is presented by means of permutation matrices. Thus for $n=5$, $\lambda=(3,2)$ there are 10 products $x_1x_2,\cdots,x_4x_5$; the characters of $\Delta(3,2)$ are the number of these products which are left invariant by the permutations of the various classes and are at once seen to be $(10,4,1,2,0,1,0)$ which checks with the result

$$q_3(s)q_2(s)=\frac{1}{5!}\{10s_1{}^5+40s_1{}^3s_2+20s_1{}^2s_3+30s_1s_2{}^2+20s_2s_3\}$$

2. $\lambda_1=n-2$, $\lambda_2=1$, $\lambda_3=1$. The cosets permute like the products $x_1{}^2x_2$. For

$n = 5$. $\Delta(3, 1^2)$ is of dimension 20 and its characters are $(20, 6, 2, 0, 0, 0, 0)$ (the permutation (12), for instance, leaving invariant the six products $x_3{}^2x_4$, $x_3x_4{}^2$, $x_4{}^2x_5$, $x_4x_5{}^2$, $x_5{}^2x_3$, $x_3x_5{}^2$); a result which checks with

$$q_3(s)q_1{}^2(s) = \frac{1}{5!}\{20s_1{}^5 + 60s_1{}^3s_2 + 40s_1{}^2s_3\}.$$

It is clear that the space spanned by the expressions $x_1x_2(x_1 + x_2)$ etc. is an invariant subspace of the carrier space of $\Delta(n-2, 1^2)$ as is also the space spanned by the expressions $x_1{}^2(x_2 + x_3 + \cdots + x_n)$ so that $\Delta(n-2, 1^2)$ is in general reducible. Similarly the basic tensors for $\Delta(n-3, 2, 1)$ are $x_1{}^2(x_2x_3)$; for $\Delta(n-3, 1^3)$ they are $x_1{}^3x_2{}^2x_3$; for $\Delta(n-4, 3, 1)$ they are $x_1{}^2(x_2x_3x_4)$; for $\Delta(n-4, 2^2)$ they are $(x_1x_2)^2(x_3x_4)$; for $\Delta(n-4, 2, 1^2)$ they are $x_1{}^3x_2{}^2(x_3x_4)$ and so on. The attempt to solve one of the main problems of the present paper; namely, the analysis of the reducible representation $\Delta(\lambda_1, \cdots, \lambda_k)$ into its irreducible components, by the geometrical method of "tensor representations" soon becomes hopelessly complicated and we shall make no use of this geometrical viewpoint.

The subgroup H whose elements are $P = P_1P_2 \cdots P_k$ may be termed the direct product of the subgroups $G_1, G_2, \cdots, G_k$ where G_j permutes only the letters in the j-th box, leaving all the other letters fixed $(j = 1, \cdots, k)$; so that G_j is of order $\lambda_j!$ We indicate the direct product relationship thus: $H = G_1 \times G_2 \times \cdots \times G_k$ and observe that if Γ_j is any representation of G_j $(j = 1, \cdots, k)$, the Kronecker product $\Gamma_1 \times \Gamma_2 \times \cdots \times \Gamma_k$ is a representation of H whose characters are the products of the corresponding characters of $\Gamma_1, \cdots, \Gamma_k$. If $\phi_j(s)$ is the characteristic of G_j associated with Γ_j the characteristic of H associated with $\Gamma_1 \times \Gamma_2 \times \cdots \Gamma_k$ is, accordingly $\prod_{j=1}^{k} \phi_j(s)$ and this furnishes a compound characteristic of G; (when the representations Γ_j $(j = 1, 2, \cdots, k)$, are all the identity representation $\phi_j(s) = q_{\lambda_j}(s)$ and we recover the compound characteristic $\prod_{j=1}^{k} q_{\lambda_j}(s)$ of G). We shall be particularly interested in the sequel in the case $k = 2$; Γ_1 will be an irreducible representation of the symmetric group on λ_1 letters and Γ_2 an irreducible representation of the symmetric group on λ_2 letters. The compound characteristic of the symmetric group on $n = \lambda_1 + \lambda_2$ letters obtained in the manner described above corresponds to a reducible representation of this group on n letters which may be termed the direct product of Γ_1 and Γ_2 (and denoted $\Gamma_1 \cdot \Gamma_2$). Our problem is the analysis of $\Gamma_1 \cdot \Gamma_2$ into its irreducible components. The dimension of the direct product $\Gamma_1 \cdot \Gamma_2$ is the product of the dimensions of Γ_1 and Γ_2 by $n! \div \lambda_1!\,\lambda_2!$, since its characteristic is $\phi_1(s)\phi_2(s)$ and the coefficient of the highest power of s_1 in this product is $\frac{1}{\lambda_1!}\,\frac{1}{\lambda_2!}$ times the product of the dimensions of Γ_1 and Γ_2.

The formula of Frobenius for the simple characteristics of the symmetric group and its modification by Schur. If we suppose the indeterminates $(s_1, \cdots, s_n)$ which occur in the expressions for the characteristics of the symmetric group to be the power sums of other indeterminates $(z_1, \cdots, z_n)$:

$$s_k = z_1^k + z_2^k + \cdots + z_n^k; \qquad (k = 1, \cdots, n)$$

the principal characteristic $q_m(s)$ of the symmetric group on $m \leq n$ letters becomes, when expressed in terms of the indeterminates $(z_1, \cdots, z_n)$, merely the complete homogeneous symmetric function $p_m(z)$ of degree m in the n variables $(z_1, \cdots, z_n)$. The first few of these functions are

$$p_0(z) = 1;\ p_1(z) = \Sigma z_1;\ p_2(z) = \Sigma z_1^2 + \Sigma z_1 z_2;\ p_3(z) = \Sigma z_1^3 + \Sigma z_1^2 z_2 + \Sigma z_1 z_2 z_3$$

and they are the coefficients of the development of the *generating* function $f(t) = \{(1 - z_1 t)(1 - z_2 t) \cdots (1 - z_n t)\}^{-1}$ in a power series in t:

$$f(t) = p_0(z) + p_1(z)t + p_2(z)t^2 + \cdots = \sum_0^\infty p_j(z) t^j.$$

But

$$\log f(t) = \sum_{p=1}^n (z_p t + \tfrac{1}{2} z_p^2 t^2 + \tfrac{1}{3} z_p^3 t^3 + \cdots) = s_1 t + \tfrac{1}{2} s_2 t^2 + \tfrac{1}{3} s_3 t^3 + \cdots$$

so that

$$\begin{aligned} f(t) &= e^{s_1 t} e^{\frac{1}{2} s_2 t^2} e^{\frac{1}{3} s_3 t^3} \cdots \\ &= \left\{ \sum_0^\infty \frac{1}{\alpha_1 !} \left(\frac{s_1}{1}\right)^{\alpha_1} t^{\alpha_1} \right\} \left\{ \sum_0^\infty \frac{1}{\alpha_2 !} \left(\frac{s_2}{2}\right)^{\alpha_2} t^{2\alpha_2} \right\} \cdots \\ &= \sum_{j=0}^\infty \left\{ \frac{1}{\alpha_1 !} \cdot \frac{1}{\alpha_2 !} \cdots \frac{1}{\alpha_j !} \left(\frac{s_1}{1}\right)^{\alpha_1} \left(\frac{s_2}{2}\right)^{\alpha_2} \cdots \left(\frac{s_j}{j}\right)^{\alpha_j} \right\} t^{\alpha_1 + 2\alpha_2 + \cdots + j\alpha_j} \end{aligned}$$

and this implies $p_j(z) = q_j(s)$ $(j = 0, 1, 2, \cdots, n)$. The homogeneous products $p_j(z) = \Sigma (z_1^{t_1} z_2^{t_2} \cdots z_n^{t_n})$, $t_1 + t_2 + \cdots + t_n = j$ are intimately connected, in a reciprocal manner, with the elementary symmetric functions:

$$\sigma_0(z) = 1;\ \sigma_1(z) = \Sigma z_1;\ \sigma_2(z) = \Sigma z_1 z_2; \cdots\ \sigma_n(z) = z_1 z_2 \cdots z_n$$

either set being expressible as polynomials with integral coefficients in the other set. In fact the generating function for the elementary symmetric functions $\sigma_k(z)$ is

$$g(t) = \prod_{p=1}^n (1 + z_p t) = \sum_0^\infty \sigma_j(z) t^j$$

and on taking logarithms we find

$$\log g(t) = s_1 t - \tfrac{1}{2}s_2 t^2 + \tfrac{1}{3}s_3 t^3 - \cdots$$

so that $\sigma_j(z) = q_j(s_1, -s_2, s_3, -s_4, \cdots) = \pi_j(s)$. In other words the alternating characteristic of the symmetric group on $j \leq n$ letters becomes, when expressed in terms of the indeterminates $(z_1, \cdots, z_n)$, simply the elementary symmetric function $\sigma_j(z)$. The two generating functions $f(t)$ and $g(t)$ are such that $g(t) = \{f(-t)\}^{-1}$ and hence $\{\sum_0^\infty \sigma_j t^j\}\{\sum_0^\infty (-1)^k p_k t^k\} = 1$ and this yields the series of relations

$$\sigma_0 p_0 = 1; \ \ \sigma_0 p_1 - \sigma_1 p_0 = 0; \ \ \sigma_0 p_2 - \sigma_1 p_1 + \sigma_2 p_0 = 0; \cdots.$$

These may be expressed by the statement that the two matrices

$$P_j = \begin{pmatrix} p_0 & p_1 & \cdots & p_{j-1} \\ & p_0 & \cdots & p_{j-2} \\ & & \cdots & \\ & & & p_0 \end{pmatrix} \quad \text{and} \quad \Sigma_j = \begin{pmatrix} \sigma_0 & -\sigma_1 & \sigma_2 & \cdots \\ & \sigma_0 & -\sigma_1 & \cdots \\ & & & \cdots \\ & & & \sigma_0 \end{pmatrix}; \qquad (j = 1, 2, \cdots)$$

are reciprocal (the elements below the diagonal in each matrix being zero). Since $p_0 = 1 = \sigma_0$ the determinant of each matrix is unity and so each element of either is a cofactor of the other; thus

$$\sigma_1 = p_1; \quad \sigma_2 = \begin{vmatrix} p_1 & p_2 \\ p_0 & p_1 \end{vmatrix}; \quad \sigma_3 = \begin{vmatrix} p_1 & p_2 & p_3 \\ p_0 & p_1 & p_2 \\ 0 & p_0 & p_1 \end{vmatrix}$$

and so in general. σ_r is an r rowed determinant whose diagonal elements are all p_1 the non-diagonal elements being obtained by increasing the suffix carried by p methodically by one as we move from each column to its right-hand neighbor and decreasing this suffix by one as we move from each column to its left-hand neighbor (it being understood that $p_{-1} = p_{-2} = \cdots = 0$). The result of this calculation needed for our immediate purpose lies on the surface: the symmetric functions $p_j(z)$ may be used instead of the elementary symmetric functions $\sigma_k(z)$ as a basis for symmetric functions. More particularly any symmetric polynomial in z with integral coefficients may be expressed as a linear combination of products of the functions p_k with integral coefficients; and if the polynomial is homogeneous of degree n the products entering the linear combination are of the type $p_{\lambda_1}(z) p_{\lambda_2}(z) \cdots p_{\lambda_k}(z)$ where

$\lambda_1 + \lambda_2 + \cdots + \lambda_k = n$. Since $p_j(z) = q_j(s)$ this furnishes the basic result: any homogeneous symmetric polynomial of degree n in the variables $(z_1, \cdots, z_n)$, with integral coefficients, is, when expressed in terms of the variables $(s_1, \cdots, s_n)$ a generalised characteristic of the symmetric group on n letters. Frobenius' essential contribution to the theory was the discovery of those particular symmetric functions of z which yield the *simple* characteristics; and to Schur we owe the recognition of the importance of a very elegant and useful expression of them (due to Jacobi) as determinants whose elements are the functions $p_j(z) = q_j(s)$, $j \leq n$.

The expression just given for $\sigma_r(z)$ as a determinant whose elements are members of the set $p_j(z)$ is merely a special case of a general reciprocal relationship between determinants whose elements are members of the set $\sigma_j(z) = \pi_j(s)$ and determinants whose elements are members of the set $p_j(z) = q_j(s)$; which merely reflects the fact that the matrices P_j and Σ_j are reciprocal. We shall need a special case of this relationship and it is convenient to derive it here. Let $(\lambda) = (\lambda_1, \cdots, \lambda_k)$ be a partition of n and consider the determinant

$$\{\lambda\} = \begin{vmatrix} p_{\lambda_1} & p_{\lambda_1+1} & \cdots & p_{\lambda_1+k-1} \\ p_{\lambda_2-1} & p_{\lambda_2} & \cdots & p_{\lambda_2+k-2} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ p_{\lambda_k-k+1} & & \cdots & p_{\lambda_k} \end{vmatrix}.$$

This determinant is a certain k rowed minor of the matrix P_{λ_1+k}; in fact the one obtained by erasing the first λ_1 columns and retaining the 1st, the $\lambda_1 - \lambda_2 + 2)$-nd, the $(\lambda_1 - \lambda_3 + 3)$-rd, $\cdots$ and the $(\lambda_1 - \lambda_k + k)$-th rows. Save for a question of sign, into which it is profitless to go since it can be settled in a trivial manner later, $\{\lambda\}$ is, therefore, equal to that minor of the reciprocal matrix Σ_{λ_1+k} which is obtained by *keeping* the first λ_1 *rows* and *omitting* the 1-st, the $(\lambda_1 - \lambda_2 - 2)$-nd, the $(\lambda_1 - \lambda_3 + 3)$-rd $\cdots$ and the $(\lambda_1 - \lambda_k + k)$-th *columns*. Since $\lambda_k > 0$, the last column of Σ_{λ_1+k} is kept and the suffix attached to the σ in the lower right-hand corner is k (for the minor has λ_1 rows and the suffix attached to the σ in the upper right-hand corner is $\lambda_1 + k - 1$ whilst the suffixes diminish methodically by one as we step from each row to its neighbor below). Counting from the last column the first column omitted is the $(\lambda_k + 1)$-st; and the suffixes of the last λ_k diagonal elements of the minor of Σ_{λ_1+k} in question all equal k; since the second column omitted is the $(\lambda_{k-1} + 2)$-nd, counting from the last, and so on, the next diagonal suffix, counting upwards to the left, is less than k by the number

of λ's that equal λ_k. Continuing in this way we see that the diagonal suffixes of the minor of Σ_{λ_1+k} constitute the partition $(\mu) = (\mu_1, \mu_2, \cdots, \mu_{\lambda_1})$ of n which is associated with the partition (λ) of n. For instance if $n = 4$ and $(\lambda) = (2, 1^2)$ so that $(\mu) = (3, 1)$ we have proved that

$$\{2, 1^2\} = \begin{vmatrix} p_2 & p_3 & p_4 \\ p_0 & p_1 & p_2 \\ 0 & p_0 & p_1 \end{vmatrix} = \pm \begin{vmatrix} -\sigma_1 & \sigma_4 \\ \sigma_0 & -\sigma_3 \end{vmatrix}.$$

The negative signs may be removed from the σ's carrying odd labels by changing the signs of all columns having a σ with an odd suffix at the bottom and following this by a change of sign of all rows having a σ with an even suffix at the end. On reflecting the σ minor about its secondary diagonal (an operation which does not affect the value of the determinant) we find

$$\{\lambda\} \equiv \begin{vmatrix} p_{\lambda_1} & \cdots & \cdot \\ \cdot & p_{\lambda_2} & \cdot \\ \cdot & \cdots & p_{\lambda_k} \end{vmatrix} = \pm \begin{vmatrix} \sigma_{\mu_1} & \cdots & \cdot \\ \cdot & \sigma_{\mu_2} & \cdot \\ \cdot & \cdots & \sigma_{\mu_j} \end{vmatrix}; \qquad (\mu_1 = k,\ j = \lambda_1)$$

where the non-diagonal elements of each determinant are filled in by increasing methodically the suffixes by one as we move to the right (and diminishing them by one as we move to the left); it being understood that a p or σ carrying a negative suffix is to be replaced by zero and that the partitions (λ) and (μ) of n are associated. That the undetermined sign is $+$, rather than $-$, is immediately evident when we recall that $p_j(\boldsymbol{z}) = q_j(\boldsymbol{s})$

$$\sigma_j(\boldsymbol{z}) = \pi_j(\boldsymbol{s}) = q_j(s_1, -s_2, s_3, -s_4, \cdots).$$

On setting $s_1 = 1$, $s_2 = s_3 = \cdots = s_n = 0$, $p_j(\boldsymbol{z}) = q_j(\boldsymbol{s})$ takes the value $1/j!$ as also does $\sigma_j(\boldsymbol{z}) = \pi_j(\boldsymbol{s})$. And on writing

$$\lambda_1 + (k-1) = l_1;\ \lambda_2 + (k-2) = l_2;\ \cdots \lambda_k = l_k$$

it is clear that the determinant $\{\lambda\}$ becomes the quotient by $l_1!\, l_2! \cdots l_k!$ of a k rowed determinant of which the element in the j-th row and p-th column is $l_j(l_j - 1) \cdots (l_j + p + 1 - k)$ there being $k - p$ factors (the elements in the k-th column being all unity). Since the element in the j-th row and the p-th column is a polynomial in l_j of degree $k - p$ (the coefficients of the polynomial being independent of the row number j and the coefficient of the highest power being unity) it is at once clear, on subtracting from each column

an appropriate linear combination of the succeeding columns, that the determinant is equivalent to the Vandermonde determinant whose j-th row is $(l_j^{k-1}, l_j^{k-2}, \cdots, l_j, 1)$. Its value is, therefore

$$\Delta(l) = \prod_{p<q} (l_p - l_q).$$

Since the partition (λ) of n was arranged in non-increasing order: $\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_k > 0$ the numbers (l) are arranged in *descending* order: $l_1 > l_2 > \cdots > l_k > 0$ and so $\Delta(l) > 0$. Hence when $s_1 = 1, s_2 = s_3 = \cdots s_n = 0$ both the determinants

$$\begin{vmatrix} p_{\lambda_1} & \cdot & \cdot & \cdot \\ \cdot & p_{\lambda_2} & & \cdot \\ \cdot & \cdot & \cdot & p_{\lambda_k} \end{vmatrix} \quad \text{and} \quad \begin{vmatrix} \sigma_{\mu_1} & \cdot & \cdot & \cdot \\ \cdot & \sigma_{\mu_2} & & \cdot \\ \cdot & \cdot & \cdot & \sigma_{\mu_j} \end{vmatrix}$$

are positive; they must accordingly be equal and not one the negative of the other. Moreover if we set $\mu_1 + (j-1) = m_1, \cdots, \mu_j = m_j$ the two numbers $\Delta(l) \div l_1!\, l_2! \cdots l_k!$ and $\Delta(m) \div m_1!\, m_2! \cdots m_j!$ are equal. Finally we remark that the theorem of the present paragraph may be stated in the following convenient manner. Denoting by $\{\lambda\}^*$ the result of changing the signs of $s_2, s_4, \cdots$ in $\{\lambda\}$ then

$$\{\lambda\} = \{\mu\}^*$$

where $\{\lambda\}$ is the determinant

$$\begin{vmatrix} q_{\lambda_1}(s) & \cdots & \cdot \\ \cdot & q_{\lambda_2}(s) & \cdot \\ \cdot & \cdots & q_{\lambda_k}(s) \end{vmatrix}$$

and (μ) is the partition of n associated with (λ).

We now proceed to the determination of the *simple* characteristics of the symmetric group on n letters (remembering that *any* homogeneous symmetric polynomial of degree n in the n indeterminates $(z_1, \cdots, z_n)$ yields, when expressed in terms of their power sums $(s_1, \cdots, s_n)$ a *generalised* characteristic of this group). What is needed is a criterion which will decide whether or not a set of generalised characteristics are simple, and this is provided as follows. Let $\boldsymbol{\chi}_p$ $(p = 1, \cdots, r)$, denote the characters of the r non-equivalent irreducible representations of the symmetric group on n letters (r being the number of partitions of n) so that $\phi_p(\boldsymbol{s}) = (\boldsymbol{s} \cdot \boldsymbol{\chi}_p)$ are the r simple characteristics. Let $\boldsymbol{t}$ be a second indeterminate and form the products

2

$\phi_p(\boldsymbol{s}) \cdot \phi_p(\boldsymbol{t})$ and then sum these products as p runs over the values $1, 2, \cdots, r$. We have

$$\phi_p(\boldsymbol{s}) = \frac{1}{n!} \sum_P \chi_p(P) s_1^{\alpha_1} \cdots s_n^{\alpha_n}$$

where the summation is over all elements P of the symmetric group ($(\alpha) = (\alpha_1, \cdots, \alpha_n)$ indicating the class to which P belongs). Similarly

$$\phi_p(\boldsymbol{t}) = \frac{1}{n!} \sum_Q \chi_p(Q) t_1^{\beta_1} \cdots t_n^{\beta_n}$$

where $(\beta_1, \cdots, \beta_n)$ indicates the class to which Q belongs. On forming the product and summing with respect to p we have a triple summation; namely with respect to P, with respect to Q, and with respect to p. Performing first the summation with respect to p we obtain zero unless Q belongs to the same class as P (owing to the orthogonality relations amongst the simple characters). For a fixed P there are $N_{(\alpha)}$ choices of Q, namely all the Q's in the same class as P, and summation with respect to Q gives $\sum_p N_{(\alpha)} \chi_p^{(\alpha)} \chi_p^{(\alpha)} = n!$ There remains only the single summation with respect to P and we find

$$\sum_p \phi_p(\boldsymbol{s}) \phi_p(\boldsymbol{t}) = \frac{1}{n!} \sum_P (s_1 t_1)^{\alpha_1} \cdots (s_n t_n)^{\alpha_n} = q_n(\boldsymbol{u})$$

where $\boldsymbol{u} = \boldsymbol{st}$ in the sense that $u_1 = s_1 t_1$, $u_2 = s_2 t_2, \cdots, u_n = s_n t_n$. The real force of this result lies in the fact that its converse is true in the following sense: suppose we have r generalised characteristics $F_p(\boldsymbol{s})$ $(p = 1, \cdots, r)$, possessing the property that $\sum_{p=1}^{r} F_p(\boldsymbol{s}) F_p(\boldsymbol{t}) = q_n(\boldsymbol{st})$; then each of these characteristics is either a simple characteristic or the negative of one. In fact $F_p(\boldsymbol{s}) = c_p^{\alpha} \phi_\alpha(\boldsymbol{s})$, where the coefficients c_p^q are integers, positive, negative or zero and so

$$\sum_{p=1}^{r} \phi_p(\boldsymbol{s}) \phi_p(\boldsymbol{t}) = q_n(\boldsymbol{st}) = \sum_{p=1}^{r} F_p(\boldsymbol{s}) F_p(\boldsymbol{t}) = \sum_{p=1}^{r} c_p^{\alpha} c_p^{\beta} \phi_\alpha(\boldsymbol{s}) \phi_\beta(\boldsymbol{t}).$$

Now the r simple characteristics are linearly independent; for a hypothecated relation $c^\alpha \phi_\alpha(\boldsymbol{s}) = 0$ would imply $c^\alpha \chi_{.\alpha} = 0$ and this would imply $c^p = 0$ $(p = 1, \cdots, r)$, owing to the orthogonality relations (1). Equating then, the coefficients of $\phi_q(\boldsymbol{s})$ on both sides of the equation just written, we obtain

$$\phi_q(\boldsymbol{t}) = \sum_{p=1}^{r} c_p{}^q c_p{}^\beta \phi_\beta(\boldsymbol{t})\,; \qquad (q = 1, 2, \cdots, r)$$

and this implies, again on account of the linear independence of the simple characteristics,

$$\sum_{p=1}^{r} c_p{}^q c_p{}^j = 0\,;\ j \neq q\,; \qquad \sum_{p=1}^{r} (c_p{}^q)^2 = 1.$$

Since the numbers $c_p{}^q$ are integers it follows from the second of these two equations that, for a fixed q, all $c_p{}^q$ vanish save one which is ± 1; and then the remaining equations show that for a fixed p all $c_p{}^q$ vanish save one which is ± 1. In other words the generalised characteristics $F_p(\boldsymbol{s})$ are merely a rearrangement of the simple characteristics $\phi_p(\boldsymbol{s})$ followed, possibly, by a change of sign of some of them.

Let $(v_1, \cdots, v_n)$ be n integers, positive or zero, no two of which are equal, supposed arranged in descending order of magnitude: $v_1 > v_2 > \cdots > v_n$, and denote by $A(v_1, \cdots, v_n)$ the n rowed determinant of which the elements in the j-th row are the v^j-th powers of the indeterminates $(z_1, \cdots, z_n)$ $(j = 1, \cdots, n)$. When $v_1 = n - 1, v_2 = n - 2, \cdots, v_n = 0$, $A(v_1, \cdots, v_n)$ is the Vandermonde determinant whose value is the difference product $\Delta = \Delta(\boldsymbol{z}) = \prod_{j<k} (z_j - z_k)$. It is clear that $A(v_1, \cdots, v_n)$ contains Δ as a factor and since both $A(v_1, \cdots, v_n)$ and Δ are alternating functions of $\boldsymbol{z}$ the quotient is symmetric and it is at once seen to be a polynomial of degree $(v_1 + v_2 + \cdots + v_n) - (\overline{n-1} + \overline{n-2} + 1 + 0)$ with integral coefficients. If, then, $(\lambda) = (\lambda_1, \cdots, \lambda_n)$ is a partition of n and we write

$$v_1 = \lambda_1 + (n - 1), \qquad v_2 = \lambda_2 + (n - 2), \cdots, v_n = \lambda_n$$

the quotient $A(v_1, \cdots, v_n) \div \Delta$ is a symmetric polynomial of degree n, with integral coefficients, in the indeterminates $(z_1, \cdots, z_n)$; it furnishes, therefore, when expressed in terms of the power sums $(s_1, \cdots, s_n)$, a generalised characteristic of the symmetric group on n letters and the basic result of Frobenius is to the effect that the characteristics obtained in this way are *simple*. Let us denote the quotient $A(v_1, \cdots, v_n) \div \Delta$ by $\{\lambda\}$; then in order to derive the result of Frobenius we have first to show that $\sum_{(\lambda)} \{\lambda\}(\boldsymbol{s})\{\lambda\}(\boldsymbol{t}) = q_n(\boldsymbol{st})$ and then that the coefficient of s_1 in $\{\lambda\}(\boldsymbol{s})$ is positive. We first remark that the relations

$$s_j = z_1{}^j + z_2{}^j + \cdots + z_n{}^j\,; \qquad t_j = y_1{}^j + y_2{}^j + \cdots + y_n{}^j$$

imply $s_j t_j = \Sigma (z_p y_q)^j$; $(p = 1, \cdots, n;\ q = 1, \cdots, n)$. Hence if we denote the n^2 quantities $z_p y_q$ by $\boldsymbol{zy}$ the relations $\boldsymbol{s} \rightarrow \boldsymbol{z}$, $\boldsymbol{t} \rightarrow \boldsymbol{y}$ imply $\boldsymbol{st} \rightarrow \boldsymbol{zy}$ and, in particular, $q_n(\boldsymbol{st}) = p_n(\boldsymbol{zy})$; so that the first part of our problem may be re-phrased as follows: we must show that $\sum_{(\lambda)} \{\lambda\}(\boldsymbol{z})\{\lambda\}(\boldsymbol{y}) = p_n(\boldsymbol{zy})$ the summation being over the r partitions of n.

To do this we first consider a determinant of order n of which the element in the i-th row and j-th column is $(a_i + b_j)^{-1}$. On subtracting the first *column* from each of the others and removing the common factors

$$(b_1 - b_2)(b_1 - b_3) \cdots (b_1 - b_n) \div (b_1 + a_1)(b_1 + a_2) \cdots (b_1 + a_n)$$

we obtain a determinant of which the elements in the i-th row are 1, $(a_i + b_2)^{-1}, \cdots, (a_i + b_n)^{-1}$. On subtracting the first *row* of this determinant from each of the others and removing the common factors

$$(a_1 - a_2)(a_1 - a_3) \cdots (a_1 - a_n) \div (a_1 + b_2)(a_1 + b_3) \cdots (a_1 + b_n)$$

we obtain a determinant of order $n - 1$ of which the element in the i-th row and j-th column is again $(a_i + b_j)^{-1}$ where now i, j run from 2 to n instead of from 1 to n as before. It follows at once that the n-th order determinant, of which the element in the i-th row and j-th column is $(a_i + b_j)^{-1}$, has the value $\Delta(\boldsymbol{a})\Delta(\boldsymbol{b}) \div \Pi(a_i + b_j)$; $(i = 1, \cdots, n,\ j = 1, 2, \cdots, n)$, where $\Delta(\boldsymbol{a})$ denotes the difference product $(a_1 - a_2) \cdots (a_{n-1} - a_n)$ (a result due to Cauchy). On writing $a_i = \alpha_i^{-1}$, $b_j = -\beta_j$ this result of Cauchy appears in the following equivalent form: the determinant of order n of which the element in the i-th row and j-th column is $(1 - \alpha_i \beta_j)^{-1}$ has the value

$$\Delta(\boldsymbol{\alpha})\Delta(\boldsymbol{\beta}) \div \Pi(1 - \alpha_i \beta_j).$$

But if $\boldsymbol{A}$ denotes the $n \times \infty$ matrix of which the elements in the i-th *row* are $(1, \alpha_i, \alpha_i^2, \cdots)$ and $\boldsymbol{B}$ the $\infty \times n$ matrix of which the elements in the j-th *column* are $(1, \beta_j, \beta_j^2, \cdots)$ the product $\boldsymbol{A} \cdot \boldsymbol{B}$ is an $n \times n$ matrix of which the element in the i-th row and j-th column is $1 + \alpha_i \beta_j + \alpha_i^2 \beta_j^2 + \cdots$ or $(1 - \alpha_i \beta_j)^{-1}$. The determinant of the product $\boldsymbol{AB}$ may be found by selecting *any* n-th order matrix from $\boldsymbol{A}$, multiplying its determinant by the determinant of the corresponding matrix from $\boldsymbol{B}$, and adding all products so obtained; that the number of products is infinite need cause no concern since $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are indeterminates and we may regard them so chosen that the components α_i, β_j

are all < 1 in numerical magnitude so that the infinite series which appear are all absolutely convergent. All determinants of order n selected from the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are of the type $A(p_1, \cdots, p_n)$ where we may, without lack of generality, agree that $p_1 > p_2 > \cdots > p_n > 0$. Hence we have

$$\sum_{(p)} A(p_1, \cdots, p_n)(\boldsymbol{\alpha}).\, A(p_1, \cdots, p_n)(\boldsymbol{\beta}) = \Delta(\boldsymbol{\alpha})\Delta(\boldsymbol{\beta}) \div \Pi(1 - \alpha_i\beta_j).$$

On setting $\boldsymbol{\beta} = \boldsymbol{\delta} t$ i. e. $\beta_1 = \delta_1 t$, $\beta_2 = \delta_2 t, \cdots, \beta_n = \delta_n t$, where t is an indeterminate, we have $A(p_1 \cdots p_n)(\boldsymbol{\beta}) = A(p_1, \cdots, p_n)(\boldsymbol{\delta})\, t^{p_1+p_2+\cdots+p_n}$,

$$\Delta(\boldsymbol{\beta}) = \Delta(\boldsymbol{\delta})\, t^{(n-1)+(n-2)+\cdots+1+0}$$

and on writing

$$p_1 - (n-1) = \lambda_1;\ p_2 - (n-2) = \lambda_2;\ \cdots;\ p_n = \lambda_n$$

we find

$$\sum_{(\lambda)} \{\lambda\}(\boldsymbol{\alpha})\{\lambda\}(\boldsymbol{\delta})\, t^{\lambda_1+\cdots+\lambda_n} = \{\Pi(1 - \alpha_i\delta_j t)\}^{-1}$$
$$= \sum_0^\infty p_n(\boldsymbol{\alpha\delta})\, t^n.$$

On equating coefficients of t^n we obtain

$$\sum_{(\lambda)} \{\lambda\}(\boldsymbol{\alpha})\{\lambda\}(\boldsymbol{\delta}) = p_n(\boldsymbol{\alpha\delta})$$

where the summation is over all partitions (λ) of n. This proves that the symmetric polynomials $\{\lambda\}(\boldsymbol{z})$, furnish, when expressed in terms of the power sums $(s_1, \cdots, s_n)$, either simple characteristics or the negatives of these; all simple characteristics being obtained in this way. To show that we have actually the simple characteristics, and not the negatives of any of them, we must show that the coefficient of $s_1{}^n$ in $\{\lambda\}(\boldsymbol{z}) > 0$. To do this we shall first derive Jacobi's expression for $\{\lambda\}(\boldsymbol{z})$ as a determinant whose elements are members of the set $p_j(\boldsymbol{z}) = q_j(\boldsymbol{s})$, $j \leq n$. Before doing this we remark that Frobenius stated his result in a slightly different form. From

$$\phi_{(\lambda)}(\boldsymbol{s}) = A(v_1, \cdots, v_n) \div \Delta(\boldsymbol{z})$$

and (3) we have

$$s^{(a)} = \sum_{(\lambda)} \chi_{(\lambda)}{}^{(a)} A(v_1, \cdots, v_n) \div \Delta(\boldsymbol{z})$$

so that $\chi_{(\lambda)}{}^{(a)}$ is the coefficient of $A(v_1, v_2, \cdots, v_n)$ in the development of

$$\Delta(\boldsymbol{z}) s^{(a)} = \{\prod_{p<q} (z_p - z_q)\} s_1{}^{a_1} s_2{}^{a_2} \cdots s_n{}^{a_n}.$$

To obtain Jacobi's expression it is necessary to point out some trivially evident properties of the homogeneous products $p_j(z_1, \cdots, z_n)$. The generating function for these products was $\{(1-z_1t)\cdots(1-z_nt)\}^{-1}$:

$$\{(1-z_1t)\cdots(1-z_nt)\}^{-1} = \sum_0^\infty p_j(z_1, z_2, \cdots, z_n)t^n.$$

It follows, on multiplication by $(1-z_nt)$, that

$$\sum_0^\infty p_j(z_1, z_2, \cdots, z_{n-1})t^j = (1-z_nt)\sum_0^\infty p_j(z_1, z_2, \cdots, z_n)t^j$$

so that $p_j(z_1, \cdots, z_{n-1}) = p_j(z_1, \cdots, z_n) - z_np_{j-1}(z_1, \cdots, z_n)$ or, equivalently, $p_j(z_1, \cdots, z_n) = z_np_{j-1}(z_1, \cdots, z_n) + p_j(z_1, \cdots, z_{n-1})$. Applying this "reduction formula" to both terms on the right-hand side we obtain

$$\begin{aligned} p_j(z_1, z_2, \cdots, z_n) &= z_n\{z_np_{j-2}(z_1, \cdots, z_n) + p_{j-1}(z_1, \cdots, z_{n-1})\} \\ &\qquad + z_{n-1}p_{j-1}(z_1, \cdots, z_{n-1}) + p_j(z_1, \cdots, z_{n-2}) \\ &= z_n{}^2p_{j-2}(z_1, \cdots, z_n) + (z_n + z_{n-1})p_{j-1}(z_1, \cdots, z_{n-1}) + p_j(z_1, \cdots, z_{n-2}) \end{aligned}$$

a relation which may be written in the form

$$\begin{aligned} p_j(z_1, \cdots, z_n) &= p_2(z_n)p_{j-2}(z_1, \cdots, z_n) \\ &\qquad + p_1(z_n, z_{n-1})p_{j-1}(z_1, \cdots, z_{n-1}) + p_j(z_1, \cdots, z_{n-2}) \end{aligned}$$

which suggests the relation

$$\begin{aligned} p_j(z_1, \cdots, z_n) &= p_m(z_n)p_{j-m}(z_1, \cdots, z_n) \\ &\qquad + p_{m-1}(z_n, z_{n-1})p_{j-m+1}(z_1, \cdots, z_{n-1}) + \cdots; \quad (m = 1, 2, \cdots). \end{aligned} \tag{8}$$

That this relation actually does hold is readily proved by induction; for assuming its validity for a stated value of m its validity for $m+1$ follows at once. Thus

$$\begin{aligned} p_j(z_1, \cdots, z_n) &= p_m(z_n)\{z_np_{j-m-1}(z_1, \cdots, z_n) + p_{j-m}(z_1, \cdots, z_{n-1})\} \\ &\qquad + p_{m-1}(z_n, z_{n-1})\{z_{n-1}p_{j-m}(z_1, \cdots, z_{n-1}) + p_{j-m+1}(z_1, \cdots, z_{n-2})\} + \cdots \\ &= p_{m+1}(z_n)p_{j-m-1}(z_1, \cdots, z_n) + p_m(z_n, z_{n-1})p_{j-m}(z_1, \cdots, z_{n-1}) + \cdots. \end{aligned}$$

Since the relation (8) is true for $m = 1$ it is true for every positive integer; it being always understood that a p with a negative label is assigned the value zero. It is also understood that all $p_q(z_1, \cdots, z_s)$ are assigned the value zero when $s < 1$.

We need one other property of the homogeneous products $p_j(z_1, \cdots, z_n)$. Writing the basic recurrence relation $p_j(z_1, \cdots, z_n) = z_np_{j-1}(z_1, \cdots, z_n) + p_j(z_1, \cdots, z_{n-1})$ in the equivalent form

$$p_{j-1}(z_1, \cdots, z_n) = p_{j-1}(z_1, \cdots, z_{n+1}) - z_{n+1}p_{j-2}(z_1, \cdots, z_{n+1})$$

we find

$$p_j(z_1, \cdots, z_n) = z_n\{p_{j-1}(z_1, \cdots, z_{n+1}) - z_{n+1}p_{j-2}(z_1, \cdots, z_{n+1})\} + p_j(z_1, \cdots, z_{n-1})$$

and, on interchanging z_n and z_{n+1} and subtracting,

$$\begin{aligned} &p_{j-1}(z_1, \cdots, z_{n-1}, z_n, z_{n+1}) \\ &\qquad = \{p_j(z_1, \cdots, z_{n-1}, z_{n+1}) - p_j(z_1, \cdots, z_{n-1}, z_n)\} \div (z_{n+1} - z_n). \end{aligned} \tag{9}$$

We are now ready to carry out Jacobi's transformation of the symmetric polynomial $A(v_1,\cdots,v_n)\div\Delta(z)$. The determinant $A(v_1,\cdots,v_n)$ has $z_j^{v_i}=p_{v_i}(z_j)$ as the element in the i-th row and j-th column. On subtracting the last column from each of the others and removing the factors

$$(z_1-z_n)(z_2-z_n)\cdots(z_{n-1}-z_n)$$

we obtain an n-th order determinant of which the element in the i-th row and j-th column is $p_{v_i-1}(z_j,z_n)$ $(j=1,\cdots,n-1)$; the element in the i-th row and n-th column remaining $p_{v_i}(z_n)$. We now subtract the $(n-1)$-st column from each of the columns which precede it and remove the factors

$$(z_1-z_{n-1})(z_2-z_{n-1})\cdots(z_{n-2}-z_{n-1})$$

obtaining a determinant of which the element in the i-th row and j-th column is $p_{v_i-2}(z_j,z_{n-1},z_n)$ $(j=1,\cdots,n-2)$, the elements in the last two columns remaining $p_{v_i-1}(z_{n-1},z_n)$ and $p_{v_i}(z_n)$. Proceeding in this way we see that $\{\lambda\}(z)\equiv A(v_1,\cdots,v_n)\div\Delta(z)$ may be expressed as a determinant of order n of which the element in the i-th row and j-th column is $p_{v_i-(n-j)}(z_j,z_{j+1},\cdots,z_n)$ $(j=1,2,\cdots,n)$, (it being always understood that the p's with negative labels vanish). Upon multiplying this determinant by unity in the form of an n-th order determinant of which the element in the i-th row and j-th column is $p_{j-i}(z_1,\cdots,z_i)$ (so that the diagonal elements are all unity whilst the elements below the diagonal vanish) and using (8) (the multiplication is done row into column as in matrix multiplication) we find that

$$A(v_1,\cdots,v_n)\div\Delta(z)$$

may be expressed as an n-th order determinant of which the element in the i-th row and j-th column is $p_{v_i-(n-j)}(z)=q_{v_i-(n-j)}(s)$. On setting

$$v_i=\lambda_i+n-i, \qquad (i=1,\cdots,n)$$

we see that $\{\lambda\}(z)$ is expressible as an n-th order determinant whose diagonal elements are $q_{\lambda_i}(s)$ the other elements in any row being obtained by methodically increasing (decreasing) the suffix carried by $q(s)$ as we move from any column to its neighbor on the right (left). If k is such that $\lambda_k>0$ whilst $\lambda_{k+1}=\lambda_{k+2}=\cdots=\lambda_n=0$ the last $n-k$ rows of our determinant have unity in the diagonal and zero's preceding the diagonal. Hence, and this is the

essential simplification, $\{\lambda\}(z)$ may be expressed as a determinant of order k of the type described above. The coefficient of $s_1{}^n$ is obtained by setting $s_1 = 1, s_2 = \cdots = s_n = 0$ and turns out to be positive (the calculation having been performed already on p. 453).

We restate the main theorem (Frobenius-Schur) of the present section as follows: Attached to each partition (λ) of n: $\lambda_1 + \lambda_2 + \cdots + \lambda_k = n$, $\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_k > 0$, is an irreducible representation $D(\lambda)$ of the symmetric group on n letters. Its characteristic is the determinant of order k

$$\{\lambda\}(z) = \phi_{(\lambda)}(s) = \begin{vmatrix} q_{\lambda_1}(s) & \cdots & \cdot \\ \cdot & q_{\lambda_2}(s) & \cdot \\ \cdot & \cdots & q_{\lambda_k}(s) \end{vmatrix} \tag{10}$$

(where the remaining elements of any row are obtained from the diagonal element by methodically increasing (decreasing) by unity the suffix carried by $q(s)$ as we move from any column to its neighbor on the right (left)). The characteristic of the irreducible representation $D(\mu)$ which is attached to the associated partition (μ) of n is

$$\phi_{(\mu)}(s) = \begin{vmatrix} q_{\mu_1}(s) & \cdots & \cdot \\ \cdot & q_{\mu_2}(s) & \cdot \\ \cdot & \cdots & q_{\mu_j}(s) \end{vmatrix} = \begin{vmatrix} \pi_{\lambda_1}(s) & \cdots & \cdot \\ \cdot & \pi_{\lambda_2}(s) & \cdot \\ \cdot & \cdots & \pi_{\lambda_k}(s) \end{vmatrix}$$

so that $\phi_{(\mu)}(s) = \phi_{(\lambda)}(s_1, -s_2, s_3, -s_4, \cdots)$. In other words, the characters of $D(\mu)$ which correspond to *even* classes are the *same* as the characters of $D(\lambda)$ whilst those which correspond to *odd* classes are the *negatives* of the characters of $D(\lambda)$; so that in constructing the character tables it is unnecessary to give the characters of $D(\mu)$ if those of $D(\lambda)$ have been already given. The common dimension of the irreducible representations $D(\lambda)$, $D(\mu)$ is

$$n!\,\Delta(l) \div l_1!\,l_2!\cdots l_k! \tag{11}$$

where

$l_1 = \lambda_1 + (k-1), l_2 = \lambda_2 + (k-2), \cdots, l_k = \lambda_k$ and $\Delta(l) = \prod_{p<q} (l_p - l_q)$.

The construction of the character tables for the various symmetric groups. From the expression (10) for $\phi_\lambda(s)$ and the relations (5) it follows at once, on applying the rule for differentiating a determinant, that $p\partial\phi_{(\lambda)}(s)/\partial s_p$ is the sum of k determinants of which the j-th differs from $\phi_{(\lambda)}(s)$ in that the suffixes of the q's in the j-th row are all diminished by p; $(p = 1, \cdots, n)$.

The suffixes of the diagonal elements of this j-th determinant, namely $(\lambda_1, \lambda_2, \cdots, \lambda_j - p, \cdots, \lambda_k)$ add up to $n - p$ but they will not, in general, constitute a partition of $n - p$ for $\lambda_j - p$ may well be negative and even if it is not the normal non-increasing order may well be destroyed. However, an interchange of two adjacent rows of our determinant, which amounts only to a change of its sign, changes two adjacent diagonal suffixes by interchanging them and at the same time decreasing the one which was originally on the right by unity and increasing the one which was originally on the left by unity. By doing this sufficiently often the sequence $(\lambda_1, \lambda_2, \cdots, \lambda_j - p, \cdots, \lambda_k)$ may be put in non-ascending order. If it then *ends* in a negative integer we discard the corresponding determinant, whose last row consists entirely of zeros; if it ends in one or more zeros we ignore *these* as the corresponding determinant has units in the diagonal places in the last one or more rows, all preceding elements in these rows being zero. We shall understand by $\{\lambda_1, \lambda_2, \cdots, \lambda_j - p, \cdots, \lambda_k\}$ the simple characteristic of the symmetric group on $n - p$ letters ($p = 1, 2, \cdots, n - 1$) corresponding to the partition of $n - p$ obtained in this way *provided the number of necessary interchanges is even* and the negative of this simple characteristic of the number of interchanges is *odd.* Since $\{\cdots a, b, \cdots\} = -\{\cdots b - 1, a + 1 \cdots\}$ it is clear that $\{\cdots a, b \cdots\} = 0$ if $b = a + 1$; similarly $\{\cdots a, b, c, d \cdots\} = 0$ if $c = a + 2$ or if $d = a + 3$ and so on. With this understanding of the symbol $\{\lambda_1, \lambda_2, \cdots, \lambda_j - p, \cdots, \lambda_k\}$ we have, then,

$$p \frac{\partial \phi_{(\lambda)}(s)}{\partial s_p} = \sum_{j=1}^{k} \{\lambda_1, \lambda_2, \cdots, \lambda_j - p, \cdots, \lambda_k\} \cdots. \tag{12}$$

On writing out $\phi_\lambda(s)$ thus:

$$\phi_{(\lambda)}(s) = \sum_{(a)} \frac{\chi_{(\lambda)}{}^{(a)}}{\alpha_1!\,\alpha_2! \cdots \alpha_n!} \left(\frac{s_1}{1}\right)^{a_1} \cdots \left(\frac{s_n}{n}\right)^{a_n}$$

we have

$$p \frac{\partial \phi_{(\lambda)}(s)}{\partial s_p} = \sum_{(a')} \frac{\chi_{(\lambda)}{}^{(a)}}{\alpha_1! \cdots \alpha'_p! \cdots \alpha_n!} \left(\frac{s_1}{1}\right)^{a_1} \cdots \left(\frac{s_p}{p}\right)^{a'_p} \cdots \left(\frac{s_n}{n}\right)^{a_n}$$

where $\alpha'_p = \alpha_p - 1$ and $(\alpha') = (\alpha_1, \cdots, \alpha'_p, \cdots, \alpha_n)$ is that class of the symmetric group on $n - p$ letters which contains one less cycle on p letters than the class (α) of the symmetric group on n letters. On equating coefficients of $s_1^{a_1} \cdots s_p^{a'_p} \cdots s_n^{a_n}$ on both sides of the equation (12) we find

$$\chi_{(\lambda)}{}^{(a)} = \sum_{j=1}^{k} \chi^{(a')}_{(\lambda_1, \ldots, \lambda_j - p, \ldots, \lambda_k)}$$

a relation which we find convenient to write in the form

$$(13) \qquad \{\lambda_1, \cdots, \lambda_k\}_{(a)} = \sum_{j=1}^{k} \{\lambda_1, \lambda_2, \cdots, \lambda_j - p, \cdots, \lambda_k\}_{(a')}.$$

This basic formula enables us to write down at once those characters of the symmetric group on n letters which correspond to a class containing at least one cycle on p letters when the characters of that class of the symmetric group on $n-p$ letters which contains one less cycle on p letters are known; $(p = 1, \cdots, n-1)$. The same formula yields directly the characters of the class containing but one cycle on n letters; since $\lambda_k \geq 1$, $\lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_k \geq 1$ we have $\lambda_1 + (k-1) \leq n$ the equality holding only when $\lambda_2 = \lambda_3 = \cdots = \lambda_k = 1$ and so $\lambda_1 - n + (k-1) \leq 0$ and this implies $\{\lambda_1 - n, \lambda_2, \cdots, \lambda_k\} = 0$ unless $\lambda_2 = \lambda_3 = \cdots = \lambda_k = 1$ since then the last term, when it is rearranged in non-increasing order, namely $\lambda_1 - n + k - 1, < 0$. The other terms $\{\lambda_1, \lambda_2 - n, \lambda_3, \cdots, \lambda_k\}$ etc., are zero for *all* partitions (λ) since $\lambda_2 - n + (k-2) < \lambda_1 - n + k - 1 < 0$ and so on. Hence the characters of the class containing but one cycle on n letters are zero unless the partition (λ) is of the type $(n-k+1, 1^{k-1})$. On subtracting n from the first number $n-k+1$ of this partition of n we obtain $\{1-k, 1^{k-1}\}$ and $k-1$ rearrangements are necessary to write this as $\{0^k\}$ which $= 1$. Since

$$n \frac{\partial \phi_{(\lambda)}(s)}{\partial s_n} = \chi_{(\lambda)}{}^{a_n=1}$$

we have

$$(14) \qquad \chi^{a_n=1}_{(n-k+1,\, 1^{k-1})} = (-1)^{k-1}; \quad \text{all other} \quad \chi_{(\lambda)}{}^{a_n=1} = 0.$$

This formula has the definite advantage, over the recurrence formula (13), that it tells us explicitly, *without referring to data concerning the symmetric group on a lesser number of letters,* the characters attached to a *particular class* of the symmetric group on n letters, namely the class containing but one cycle on n letters. The formula (11) of Frobenius giving the dimension of $D(\lambda)$, or, equivalently, the character attached to the unit class, has a similar advantage. We may combine our recurrence formula with the dimension formula of Frobenius to determine *directly* characters of classes containing one or more unary cycles. E. g., suppose we wish to calculate the characters of the symmetric group on $n = 20$ letters corresponding to the class containing $\alpha_1 = 12$ unary and $\alpha_8 = 1$ cycle on 8 letters. We shall illustrate by considering the representation $D(9, 6, 3, 2)$. Applying our recurrence formula with $p = 8$ we obtain

$$\{9, 6, 3, 2\}_{\alpha} = \{1, 6, 3, 2\}_{\alpha'} + \{9, -2, 3, 2\}_{\alpha'} + \{9, 6, -5, 2\}_{\alpha'} + \{9, 6, 3, -6\}_{\alpha'};$$

of the four terms on the right the first, third and fourth vanish; the first because $3 = 1 + 2$; the third because $-5 + 1 < 0$ and the fourth because $-6 < 0$. There remains $\{9, -2, 3, 2\}_{\alpha'} = -\{9, 2, -1, 2\}_{\alpha'} = \{9, 2, 1\}_{\alpha'}$ and since α' is the unit class the dimension formula of Frobenius yields, since

$$(l_1, l_2, l_3) = (11, 3, 1), \qquad \frac{12\,!}{11\,!\,3\,!\,1\,!}(8)(10)(2) = 320.$$

Similar, although not quite such convenient, formulae may be found for the characters of a class containing only cycles of the same length. E. g., let $n = 2m$ and consider the class containing m binary cycles. The characters of this class are found by setting $s_2 = 1$, $s_1 = s_3 = \cdots = s_n = 0$ in the expressions for the simple characteristics; it being clear that then $q_j = 0$ if j is odd whilst $q_{2p} = 1/2^p.\,p\,!$. Thus, for $n = 12$, the character of the class $\alpha_2 = 6$ of $D(5, 4, 2, 1)$ is

$$2^6.\,6\,! \begin{vmatrix} 0 & (2^3.\,3\,!)^{-1} & 0 & (2^4.\,4\,!)^{-1} \\ 0 & (2^2.\,2\,!)^{-1} & 0 & (2^3.\,3\,!)^{-1} \\ 1 & 0 & 2^{-1} & 0 \\ 0 & 0 & 1 & 0 \end{vmatrix} = 6\,!\,\{(4\,!\,2\,!)^{-1} - (3\,!)^{-2}\} = -5.$$

Similarly the character of the class $\alpha_3 = 4$ of $D(6, 3^2)$ is

$$3^4.\,4\,! \begin{vmatrix} (3^2.\,2\,!)^{-1} & 0 & 0 \\ 0 & 3^{-1} & 0 \\ 0 & 0 & 3^{-1} \end{vmatrix} = 12$$

whilst the character of the class $\alpha_4 = 3$ of $D(7, 2^2, 1)$ is

$$4^3.\,3\,! \begin{vmatrix} 0 & (4^2.\,2\,!)^{-1} & 0 & 0 \\ 0 & 0 & 0 & 4^{-1} \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{vmatrix} = -3.$$

Two more examples will suffice; suppose for $n = 15$ we wish the character of the class $\alpha_5 = 3$ for the representation $D(5, 4, 3, 2, 1)$. On setting $s_5 = 1$ $s_1 = s_2 = \cdots = s_{15} = 0$ all the q_j vanish save those for which j is a multiple of 5 and q_{5p} takes the value $(5^p.\,p\,!)^{-1}$. Then the desired character =

$$5^3 . 3 ! \begin{vmatrix} 5^{-1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{vmatrix} = -150.$$

If we wish, for a final example, to obtain for $n = 12$ the character of the class $\alpha_1 = 1$, $\alpha_2 = 2$, $\alpha_3 = 3$ for the representation $D(7, 1^5)$ we may proceed as follows. On applying our recurrence formula twice, first with $p = 1$ and then with $p = 2$, we find

$$\begin{aligned} \{7, 1^5\}_\alpha &= \{6, 1^5\}_{\alpha'} + \{7, 1^4\}_{\alpha'} \\ &= \{4, 1^5\}_{\alpha''} - \{6, 1^3\}_{\alpha''} + \{5, 1^4\}_{\alpha''} - \{7, 1^2\}_{\alpha''} \end{aligned}$$

where α'' is the class, of the symmetric group on 9 letters, consisting of permutations each of which has three ternary cycles. Since this class is positive $\{4, 1^5\}_{\alpha''} = \{6, 1^3\}_{\alpha''}$ and we have merely to calculate $\{5, 1^4\}_{\alpha''}$ and $\{7, 1^2\}_{\alpha''}$. We find

$$\{5, 1^4\}_{\alpha''} = 3^3 . 3 ! \begin{vmatrix} 0 & (3^2 . 2 !)^{-1} & 0 & 0 & (3^3 . 3 !)^{-1} \\ 1 & 0 & 0 & 3^{-1} & 0 \\ 0 & 1 & 0 & 0 & 3^{-1} \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{vmatrix} = -2$$

$$\{7, 1^2\}_{\alpha''} = 3^3 . 3 ! \begin{vmatrix} 0 & 0 & (3^3 . 3 !)^{-1} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} = 1$$

so that the desired character is -3. The most trivial example of this method furnishes directly the characters of the class $\alpha_1 = 1$, $\alpha_{n-1} = 1$ of the symmetric group on n letters. All characters are zero save that of the identical representation and those associated with the partitions

$$\lambda_1 = n - k, \qquad \lambda_2 = 2, \qquad \lambda_3 = \cdot \cdot \cdot = \lambda_k = 1$$

and these have the value $(-1)^{k-1}$.

The character tables for the various symmetric groups from $n = 2$ to $n = 10$, inclusive, are given in the paper numbered **(15)** in the list of references. The character table for $n = 11$ is given in the paper numbered **(16)**; (anyone

using this table should note that the characters of $(5, 3, 1^3)$ are given with the wrong signs for the odd classes). And the character tables for $n = 12$ and $n = 13$ are given in the paper numbered (**17**).

We may derive by the method just described explicit formulae for the characters of those classes of the symmetric group on n letters which consist of $\alpha_1 = n - p$ unary cycles and $\alpha_p = 1$ cycle on p letters; $(p = 2, 3, 4, \cdots)$. These formulae were given by Frobenius (**4**) for $p = 2$ (the transposition class) and (**5**) for $p = 3, 4$. Since they are of importance in the physical applications we give their derivation here. We first remark that a partition (λ) of n may be conveniently specified as follows: draw the principal diagonal of the diagram of the partition (i. e. the diagonal starting at the upper left-hand corner) and suppose it strikes s columns. Denote by $b_1 > b_2 > \cdots > b_s \geqq 0$ the number of dots to the right of the diagonal in the rows $1, 2, \cdots, s$, respectively, and by $a_1 > a_2 > \cdots > a_s \geqq 0$ the number of dots below the diagonal in the columns $1, \cdots, s$ respectively. Then the partition is described by $\boldsymbol{b} = (b_1, \cdots, b_s)$ and $\boldsymbol{a} = (a_1, \cdots, a_s)$ it being clear that the partition is self-associated when and only when $\boldsymbol{a} = \boldsymbol{b}$. The number of dots in the first row and column together $= b_1 + a_1 + 1$; when these are deleted the number of dots in the new first row and column $= b_2 + a_2 + 1$. Proceeding in this way we have $n = \sum_{j=1}^{s} (b_j + a_j + 1)$. It is clear from the definition that $b_j = \lambda_j - j$ $(j = 1, \cdots, s)$, whilst the differences $p - \lambda_{p+1}$ $(p = s, \cdots, k-1)$, satisfy the inequalities

$$0 \leqq s - \lambda_{s+1} < s + 1 - \lambda_{s+2} < \cdots < k - 1 - \lambda_k < k - 1.$$

Hence they are the complementary set to the set $a_1 > a_2 > \cdots > a_s$ in the set $0, 1, \cdots, k-1$. In fact $\lambda_k > 0$ shows that $a_1 = k - 1$ is not in the set; if $\lambda_k > 1$, $a_2 = k - 2$ is not in the set and so on. The following will serve as illustrations of the definitions of $\boldsymbol{b}$ and $\boldsymbol{a}$:

$$(\lambda) = (3, 2^2, 1^2)\,;\ s = 2;\ \boldsymbol{b} = (2, 0)\,;\quad \boldsymbol{a} = (4, 1)$$
$$(\lambda) = (4, 2, 1^2)\,;\ s = 2;\ \boldsymbol{b} = (3, 0)\,;\quad \boldsymbol{a} = (3, 0)$$
$$(\lambda) = (4^3, 1)\,;\quad s = 3;\ \boldsymbol{b} = (3, 2, 1)\,;\ \boldsymbol{a} = (3, 1, 0).$$

We denote, for convenience, by $\chi_{(\lambda)}(p)$ the characters of the class $\alpha_1 = n - p$, $\alpha_p = 1$ so that, for instance, $\chi_{(\lambda)}(2)$ are the characters of the transposition class whilst $\chi_{(\lambda)}(1)$ are the characters of the unit class (i. e. the dimensions

of the various irreducible representations). Our object is to obtain for $\chi_{(\lambda)}(p)$ $(p=2,3,4,\cdots)$, an explicit formula analogous to (11) which furnishes $\chi_{(\lambda)}(1)$. The recurrence formula (13) tells us that $\chi_{(\lambda)}(p)$ is the sum of the dimensions of the irreducible representations

$$D(\lambda_1,\lambda_2,\cdots,\lambda_j-p,\cdots,\lambda_k) \qquad (j=1,\cdots,k),$$

of the symmetric group on $n-p$ letters (where we follow the previously agreed on convention for the restoration of the normal non-increasing order of the $(\lambda_1,\lambda_2,\cdots)$ when this has been destroyed by the subtraction of p). Writing, as before,

$$l_1=\lambda_1+(k-1), \qquad l_2=\lambda_2+(k-2),\cdots,l_k=\lambda_k$$

the dimension of $D(\lambda_1-p,\lambda_2,\cdots,\lambda_k)$ is

$$(n-p)!\,(l_1-p-l_2)\cdots(l_1-p-l_k)(l_2-l_3)\cdots(l_{k-1}-l_k) \div (l_1-p)!\,l_2!\cdots l_k!$$

and there are similar expressions of the dimensions of the other irreducible representations. On dividing through by

$$\chi_{(\lambda)}(1)=n!\,(l_1-l_2)\cdots(l_{k-1}-l_k)\div l_1!\cdots l_k!$$

the quotient $\chi_{(\lambda)}(p)\div\chi_{(\lambda)}(1)$ appears as a sum of k terms of which the first is

$$l_1(l_1-1)\cdots(l_1-p+1)(l_1-p-l_2)\cdots(l_1-p-l_k) \div n(n-1)\cdots(n-p+1)(l_1-l_2)\cdots(l_1-l_k).$$

If we write $f(x)\equiv(x-l_1)\cdots(x-l_k)$ this may be written as the quotient of $l_1(l_1-1)\cdots(l_1-p+1)f(l_1-p)$ by $-pn(n-1)\cdots(n-p+1)f'(l_1)$ where f' indicates the derivative of f; hence

$$\frac{\chi_{(\lambda)}(p)}{\chi_{(\lambda)}(1)}=\frac{-1}{pn(n-1)\cdots(n-p+1)}\sum_{j=1}^{k}\frac{l_j(l_j-1)\cdots(l_j-p+1)f(l_j-p)}{f'(l_j)}.$$

Now the analysis of the function $x(x-1)\cdots(x-p+1)f(x-p)\div f(x)$ into simple fractions yields a polynomial in x plus terms $A_j\div(x-l_j)$ where $A_j=l_j(l_j-1)\cdots(l_j-p+1)f(l_j-p)\div f'(l_j)$ so that

$$\chi_{(\lambda)}(p)\div\chi_{(\lambda)}(1)=-\left(\sum_{j=1}^{k}A_j\right)\div pn(n-1)\cdots(n-p+1).$$

On writing $(x-l_j)^{-1}=(1/x)+(l_j/x^2)+\cdots$ it is clear that $\sum_{j=1}^{k}A_j$ is the coefficient of $(1/x)$ in the development of

$$x(x-1)\cdots(x-p+1)f(x-p)\div f(x)$$

as a series of *descending* powers of x. The zeros of $f(x)$ are the k numbers $(l_1,\cdots,l_k)$ so that, if $y=x-k$, the zeros of $f(y+k)$ are the k numbers $l_1-k,\cdots,l_k-k$ i. e. the k numbers λ_j-j $(j=1,\cdots,k)$. Of these the first s are the numbers $(b_1,\cdots,b_s)$ whilst the remaining $k-s$ are the negatives of α_j+1 where the two sets $\boldsymbol{a}=(a_1,\cdots,a_s)$ and $\boldsymbol{\alpha}=(\alpha_{s+1},\cdots,\alpha_k)$ together form the set $0,1,\cdots,(k-1)$. Hence

$$\begin{aligned} f(y+k) &= \prod_{j=1}^{s}(y-b_j)\cdot\prod_{s+1}^{k}(y+\alpha_h+1) \\ &= \prod_{j=1}^{s}\{(y-b_j)/(y+a_j+1)\}.\,(y+1)\cdots(y+k). \end{aligned}$$

It will be convenient to denote the function

$$(y-b_1)\cdots(y-b_s)/(y+a_1+1)\cdots(y+a_s+1)$$

by $F(y)$ and then $f(x)=f(y+k)=F(y)(y+1)\cdots(y+k)$. The desired sum $\sum_{j=1}^{k}A_j$, being the coefficient of $1/x$ in the development of

$$x(x-1)\cdots(x-p+1)f(x-p)\div f(x)$$

in a series of descending powers of x, is, equivalently, the coefficient of $1/y$ in the development of this same function in a descending series of powers of y. But

$$\begin{aligned} &x(x-1)\cdots(x-p+1)f(x-p)\div f(x) \\ &\quad=(y+k)\cdots(y+k+1-p)\cdot(y+k-p)\cdots(y+1-p)F(y-p) \\ &\qquad\div(y+1)\cdots(y+k)F(y) \\ &\quad=y(y-1)\cdots(y-p+1)F(y-p)\div F(y) \end{aligned}$$

and we have merely to seek the coefficient of $(1/y)$ in the development of this function. An application of Taylor's expansion yields

$$\begin{aligned} F(y-p)\div F(y) = 1 &- pF'(y)/F(y)+p^2F''(y)/2\,!\,F(y) \\ &- p^3F'''(y)/3\,!\,F(y)+\cdots; \end{aligned}$$

on taking the logarithmic derivative of

$$F(y)=\prod_{j=1}^{s}\{(y-b_j)/(y+a_j+1)\}$$

we find

$$F'(y)/F(y) = \sum_{j=1}^{s} \{[1/(y-b_j)] - [1/(y+a_j+1)]\}$$
$$= (n/y^2) + (c_3/y^3) + (c_4/y^4) + \cdots$$

where

$$c_3 = \sum_{j=1}^{s} \{b_j^2 - (a_j+1)^2\};\ c_4 = \sum_{j=1}^{s} \{b_j^3 + (a_j+1)^3\};\ c_5 = \sum_{j=1}^{s} \{b_j^4 - (a_j+1)^4\} \cdots$$

(we have availed ourselves of the relation $\sum_{j=1}^{s} \{b_j + (a_j+1)\} = n$). On successive differentiation of this relation we find

$$F''(y)/F(y) = \{F'(y)/F(y)\}^2 + \{F'(y)/F(y)\}'$$
$$= (-2n/y^3) + \{(n^2-3c_3)/y^4\} + \{(2nc_3-4c_4)/y^5\} + \cdots$$
$$F'''(y)/F(y) = \{F''(y)/F(y)\}\{F'(y)/F(y)\} + \{F''(y)/F(y)\}'$$
$$= (6n/y^4) + \{(12c_3-6n^2)/y^5\} + \cdots$$
$$F''''(y)/F(y) = \{F'''(y)/F(y)\}\{F'(y)/F(y)\} + \{F'''(y)/F(y)\}'$$
$$= (-24n/y^5) + \cdots.$$

Hence

$$F(y-p) \div -pF(y) = -(1/p) + (n/y^2) + (pn+c_3)/y^3$$
$$+ (2c_4 + 3pc_3 + 2np^2 - pn^2)/y^4$$
$$+ \{c_5 + 2pc_4 + p(2p-n)c_3 + np^2(p-n)\}/y^5 \cdots.$$

This has to be multiplied by $y(y-1)\cdots(y-p+1)$ and the coefficient of y^{-1} in the product then determined; equivalently we may multiply by $(y-1)\cdots(y-p+1)$ and determine the coefficient of y^{-2}. This coefficient yields, when divided by $n(n-1)\cdots(n-p+1)$ the desired quantity $\chi_{(\lambda)}(p) \div \chi_{(\lambda)}(1)$. We carry out the calculation for $p = 2, 3, 4$.

$$p = 2; \qquad \chi_{(\lambda)}(2) \div \chi_{(\lambda)}(1) = (n+c_3) \div n(n-1).$$

Since

$$c_3 = \sum_{j=1}^{s} \{b_j^2 - (a_j+1)^2\}, \qquad n = \sum_{j=1}^{s} \{b_j + (a_j+1)\}$$
$$c_3 + n = \sum_{j=1}^{s} \{b_j(b_j+1) - a_j(a_j+1)\}$$

so that

$$(15) \quad \chi_{(\lambda)}(2) \div \chi_{(\lambda)}(1) = \left[\sum_{j=1}^{s} \{b_j(b_j+1) - a_j(a_j+1)\}\right] \div n(n-1)$$

$p = 3$; here we must multiply by $(y-1)(y-2)$ and the coefficient of y^{-2} is

$$(2c_4+3c_3-3n^2+4n)/2=\tfrac{1}{2}[\sum_{j=1}^{s}\{(2b_j^3+3b_j^2+b_j) \\ +2(a_j+1)^3-3(a_j+1)^2+(a_j+1)\}-3n(n-1)].$$

Hence

$$(16)\quad \chi_{(\lambda)}(3)\div\chi_{(\lambda)}(1)=[\sum_{j=1}^{s}\{b_j(b_j+1)(2b_j+1) \\ +a_j(a_j+1)(2a_j+1)\}-3n(n-1)]\div 2n(n-1)(n-2)$$

$p=4$; here we must multiply by

$$(y-1)(y-2)(y-3)=y^3-6y^2+11y-6$$

and the coefficient of y^{-2} is $c_5+2c_4+c_3-2(2n-3)(c_3+n)$

$$=\sum_{j=1}^{s}[\{b_j^2(b_j+1)^2-a_j^2(a_j+1)^2\}-2(2n-3)\{b_j(b_j+1)-a_j(a_j+1)\}]$$

so that

$$(17)\quad \chi_{(\lambda)}(4)\div\chi_{(\lambda)}(1)=[\sum_{j=1}^{s}\{b_j^2(b_j+1)^2-a_j^2(a_j+1)^2\} \\ -2(2n-3)\{b_j(b_j+1)-a_j(a_j+1)\}]\div n(n-1)(n-2)(n-3).$$

For higher values of p it is more serviceable to use the recurrence formula (13) as the expressions deduced by the manner described above become too complicated. The formula (15) of Frobenius may be readily transformed into an equivalent formula due to Hund (see reference (**23**)). We have $b_j=\lambda_j-j$, $(j=1,\cdots,s)$, $\alpha_j+1=j-\lambda_j$, $(j=s+1,\cdots,k)$ where $\boldsymbol{a}=(a_1,\cdots,a_s)$ and $\boldsymbol{\alpha}=(\alpha_{s+1},\cdots,\alpha_k)$ together form the set $(0,\cdots,k-1)$; so that

$$\sum_1^s a_j(a_j+1)=\sum_0^{k-1}p(p+1)-\sum_{s+1}^{k}\alpha_j(\alpha_j+1) \\ =\sum_1^k p(p-1)-\sum_{s+1}^{k}(\lambda_j-j)(\lambda_j-j+1).$$

Hence

$$\sum_1^s\{b_j(b_j+1)-a_j(a_j+1)\}=\sum_1^k(\lambda_j-j)(\lambda_j-j+1)-\sum_1^k p(p-1) \\ =\sum_1^k\lambda_j(\lambda_j-2j+1)$$

so that

$$\chi_{(\lambda)}(2)\div\chi_{\lambda}(1)=\sum_1^k\lambda_j(\lambda_j-2j+1)\div n(n-1)$$

which is Hund's formula.

The analysis of the reducible representations $\Delta(\lambda_1,\cdots,\lambda_k)$ into irreducible components. The characteristic of $\Delta(\lambda_1,\cdots,\lambda_k)$ is $q_{\lambda_1}(\boldsymbol{s})\cdots q_{\lambda_k}(\boldsymbol{s})$ whilst that of the irreducible representation $D(\lambda_1,\cdots,\lambda_k)$ is

$$\{\lambda\}(s) = \begin{vmatrix} q_{\lambda_1}(s) & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & q_{\lambda_k}(s) \end{vmatrix};$$

the problem confronting us is that of writing $q_{\lambda_1}(s)\cdots q_{\lambda_k}(s)$ as a linear combination, with positive or zero integral coefficients, of the various simple characteristics $\{\lambda\}(s)$. When $k=2$ the solution is trivially evident:

$$q_{\lambda_1}(s)\,q_{\lambda_2}(s) = \begin{vmatrix} q_{\lambda_1}(s) & q_{\lambda_1+1}(s) \\ q_{\lambda_2-1}(s) & q_{\lambda_2}(s) \end{vmatrix} + \begin{vmatrix} q_{\lambda_1+1}(s) & q_{\lambda_1+2}(s) \\ q_{\lambda_2-2}(s) & q_{\lambda_2-1}(s) \end{vmatrix} + \cdots$$

so that

$$(18)\quad \Delta(\lambda_1,\lambda_2) = D(\lambda_1,\lambda_2) + D(\lambda_1+1,\lambda_2-1) + D(\lambda_1+2,\lambda_2-2) + \cdots + D(n).$$

For $k>2$ the problem may be solved as follows (we illustrate by considering the case $k=3$). Let x_j be an operator whose effect is to replace $q_{\lambda_j}(s)$ by $q_{\lambda_j+1}(s)$: $x_j q_{\lambda_j}(s) = q_{\lambda_j+1}(s)$ $(j=1,2,3)$. Then the determinant

$$\{\lambda\}(s) = \begin{vmatrix} q_{\lambda_1}(s) & q_{\lambda_1+1}(s) & q_{\lambda_1+2}(s) \\ q_{\lambda_2-1}(s) & q_{\lambda_2}(s) & q_{\lambda_2+1}(s) \\ q_{\lambda_3-2}(s) & q_{\lambda_3-1}(s) & q_{\lambda_3}(s) \end{vmatrix}$$

may be expressed as the result of operating with

$$\begin{vmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{vmatrix}$$

upon the simple product $q_{\lambda_1}(s)q_{\lambda_2-1}(s)q_{\lambda_3-2}(s)$ and since the operators x_j operate on different symbols, x_j operating on q_{λ_j}, they are commutative so that we may apply the ordinary rules of commutative algebra. Thus

$$\{\lambda\}(s) = (x_2-x_1)(x_3-x_1)(x_3-x_2)\,q_{\lambda_1}(s)\,q_{\lambda_2-1}(s)\,q_{\lambda_3-2}(s)$$

and so

$$q_{\lambda_1}(s)\,q_{\lambda_2-1}(s)\,q_{\lambda_3-2}(s) = (x_2-x_1)^{-1}(x_3-x_1)^{-1}(x_3-x_2)^{-1}\{\lambda\}(s).$$

We write now $\xi_j = x_j^{-1}$ $(j=1,2,3)$, so that ξ_j operates on λ_j so as to *decrease* it by unity: $\xi_j q_{\lambda_j} = q_{\lambda_j-1}$. Then

$$(x_2-x_1)^{-1} = \xi_2(1-\xi_2x_1)^{-1}; \qquad (x_3-x_1)^{-1} = \xi_3(1-\xi_3x_1)^{-1};$$
$$(x_3-x_2)^{-1} = \xi_3(1-\xi_3x_2)^{-1}$$

and so

$$\begin{aligned} q_{\lambda_1}(s)\,q_{\lambda_2}(s)\,q_{\lambda_3}(s) &= x_2x_3{}^2 q_{\lambda_1}(s)\,q_{\lambda_2-1}(s)\,q_{\lambda_3-2}(s) \\ &= (1-\xi_2x_1)^{-1}(1-\xi_3x_1)^{-1}(1-\xi_3x_2)^{-1}\{\lambda\}(s). \end{aligned}$$

The product

$$(1-\xi_3x_1)^{-1}(1-\xi_3x_2)^{-1} = 1 + p_1(x_1, x_2)\xi_3 + p_2(x_1, x_2)\xi_3{}^2 + \cdots$$

and the series may be stopped at $p_{\lambda_3}(x_1, x_2)\xi_3{}^{\lambda_3}$ since each $\{\lambda\}$ with a negative number at the end vanishes. Then this must be multiplied by

$$(1-\xi_2x_1)^{-1} = 1 + p_1(x_1)\xi_2 + \cdots$$

and this series may be stopped, for each $\{\lambda'\}$, at $p_{\lambda'_2+1}(x_1)\xi_2{}^{\lambda'_2+1}$ since each $\{\lambda\}$ whose next to last member < -1 vanishes. It is clear then that $\Delta(\lambda)$ contains $D(\lambda)$ once and contains no $D(\lambda')$ for which $(\lambda) > (\lambda')$; it being understood that $(\lambda) > (\lambda')$ when the *first* non-vanishing member of the set $\lambda_j - \lambda'_j$ is positive $(j = 1, 2, \cdots)$. The argument is evidently perfectly general; thus for $k = 4$ we first operate with $1 + p_1(x_1, x_2, x_3)\xi_4 + p_2(x_1, x_2, x_3)\xi_4{}^2 + \cdots$ on $\{\lambda\}$; then follow this by $1 + p_1(x_1, x_2)\xi_3 + p_2(x_1, x_2)\xi_3{}^2 + \cdots$ and finally by $1 + p_1(x_1)\xi_2 + p_2(x_1)\xi_2{}^2 + \cdots$.

The following example will sufficiently illustrate the method: consider $\Delta(4, 2^2)$. Applying $1 + p_1(x_1, x_2)\xi_3 + p_2(x_1, x_2)\xi_3{}^2$ to $\{4, 2, 2\}$ we obtain

$$\{4, 2^2\} + \{5, 2, 1\} + \{4, 3, 1\} + \{6, 2\} + \{5, 3\} + \{4, 4\}.$$

Applying $1 + p_1(x_1)\xi_2 + \cdots$ to each of these we obtain in turn

$$\begin{aligned} &\{4, 2, 2\} + \{6, 0, 2\} + \{7, -1, 2\};\ \{5, 2, 1\} + \{6, 1^2\} + \{8, -1, 1\}; \\ &\quad \{4, 3, 1\} + \{5, 2, 1\} + \{6, 1^2\} + \{8, -1, 1\};\ \{6, 2\} + \{7, 1\} + \{8\}; \\ &\quad \{5, 3\} + \{6, 2\} + \{7, 1\} + \{8\};\ \{4, 4\} + \{5, 3\} + \{6, 2\} + \{7, 1\} + \{8\} \end{aligned}$$

and adding up we find

$$\begin{aligned} \Delta(4, 2^2) = D(4, 2^2) &+ D(4, 3, 1) + D(4^2) + 2D(5, 2, 1,) + 2D(5, 3) \\ &+ D(6, 1^2) + 3D(6, 2) + 2D(7, 1) + D(8). \end{aligned}$$

When there are but three elements $(\lambda_1, \lambda_2, \lambda_3)$ in the partition the theory just given leads to the following convenient formula. Denoting by $\overline{12}$ the operation $1 + x_1\xi_2 + x_1{}^2\xi_2{}^2 + \cdots$ we have to apply to $(\lambda_1, \lambda_2, \lambda_3)$ the operator

$$\begin{aligned} \overline{12} + (x_2 + 2x_1 \cdot \overline{12})\xi_3 &+ (x_2{}^2 + 2x_1x_2 + 3x_1{}^2 \cdot \overline{12})\xi_3{}^2 \\ &+ (x_2{}^3 + 2x_1x_2{}^2 + 3x_1{}^2x_2 + x_1{}^3 \cdot \overline{12})\xi_3{}^3 + \cdots. \end{aligned}$$

For example let us consider the analysis of $\Delta(4^2, 2)$; the application of $\overline{12}$ gives

$$\{4^2,2\}+\{5,3,2\}+\{6,2^2\}+\{7,1,2\}+\{8,0,2\}+\{9,-1,2\}$$
$$=\{4^2,2\}+\{5,3,2\}+\{6,2^2\}-\{8,1^2\}-\{9,1\};$$

$x_2\xi_3$ yields $\{4,5,1\}=0$; $2x_1\cdot\overline{12}\cdot\xi_3$ yields

$$2[\{5,4,1\}+\{6,3,1\}+\{7,2,1\}+\{8,1^2\}-\{10\}]$$

$(x_2{}^2+2x_1x_2)\xi_3{}^2$ yields

$$\{4,6\}+2\{5^2\}=\{5^2\}$$

whilst, finally, $3x_1{}^2\cdot\overline{12}\cdot\xi^2$ yields

$$3[\{6,4\}+\{7,3\}+\{8,2\}+\{9,1\}+\{10\}].$$

Combining these results we obtain

$$\begin{aligned}\Delta(4^2,2)=D(10)&+2D(9,1)+3D(8,2)+D(8,1^2)\\&+3D(7,3)+2D(7,2,1)+3D(6,4)+2D(6,3,1)\\&+D(6,2^2)+D(5^2)+2D(5,4,1)+D(5,3,2)+D(4^2,2).\end{aligned}$$

Whilst the formula just given may be regarded as a complete *theoretical* solution of the problem of analysing the reducible representation $\Delta(\lambda)$ into its irreducible components it becomes very tedious when k, the number of elements in the partition (λ), ≥ 4. Fortunately the necessary information, up to $n=11$, is available in tables prepared by Kostka.[1] This writer was interested in the general question of symmetric functions and in the course of his investigation took up the question of expressing a product $\sigma_{\lambda_1}(z)\cdots\sigma_{\lambda_k}(z)$ of the elementary symmetric functions of n variables $(z_1,\cdots,z_n)$ as a linear combination of determinants

$$\begin{vmatrix}\sigma_{n_1}(z) & \cdot & \cdots & \sigma_{n_1+j-1}(z)\\ \sigma_{n_2-1}(z) & \sigma_{n_2}(z) & \cdots & \sigma_{n_2+j-2}(z)\\ \cdot & \cdot & \cdots & \sigma_{n_j}(z)\end{vmatrix}; \qquad n_1+n_2+\cdots+n_j=n.$$

Since $\sigma_m(z)=\pi_m(s)=q_m(s_1,-s_2,s_3,-s_4,\cdots)$, $m\leq n$, it follows that such an analysis of the product $\sigma_{\lambda_1}(z)\cdots\sigma_{\lambda_k}(z)$ furnishes the analysis of the product $q_{\lambda_1}(s)\cdots q_{\lambda_k}(s)$ as a linear combination of the simple characteristics $\{n_1,\cdots,n_j\}$; or, equivalently, of the reducible representations $\Delta(\lambda_1,\cdots,\lambda_k)$ as a linear combination of the irreducible representations

[1] The fact that these tables, up to $n=8$, were published in 1882 long before the representation theory of the symmetric group, and its applications, were dreamed of, recalls to mind the verse in Ecclesiastes: "Nothing under the sun is new, neither is any man able to say: Behold this is new: for it hath already gone before in the ages that were before us."

$D(n_1, \cdots, n_j)$. The paper numbered (**18**) in the list of references gives the tables for $2 \leq n \leq 8$; that numbered (**19**) the table for $n = 9$ and that numbered (**20**), which is inaccessible to us, the tables for $10 \leq n \leq 11$. In using these tables note that the $\Delta(\lambda)$ appear at the bottom (the symbol K being used instead of Δ and the partition being indicated by a suffix, a non-decreasing rather than a non-increasing order being adopted: thus $\Delta(3^2, 2)$ appears as K_{23^2}); the $D(\lambda)$ appear on the right (the symbol C being used instead of D, and the partition being indicated by a suffix, the normal non-increasing order being used: thus $D(4, 2^2)$ appears as C_{42^2}). The tables are square but the part above the main diagonal serves another purpose and is not used in the problem that concerns us. The following method of deriving Kostka's tables (or, more particularly, that part of them which is effective for our problem) was suggested by Littlewood and Richardson (**13**), on the assumption that the character table of the symmetric group in question is at hand. The product $q_{\lambda_1}(s) \cdots q_{\lambda_k}(s)$, which is the characteristic of $\Delta(\lambda_1, \cdots, \lambda_k)$ is a linear combination, $c_{(\alpha)}{}^{(\lambda)} s^{(\alpha)}$, say, of the quantities $s^{(\alpha)} = s_1{}^{\alpha_1} s_2{}^{\alpha_2} \cdots s_n{}^{\alpha_n}$ and hence, by (3), a linear combination, $\sum_{(\beta)} c_{(\alpha)}{}^{(\lambda)} \chi_{(\beta)}{}^{(\alpha)} \phi_{(\beta)}(s)$ of the simple characteristics of the symmetric group on n letters. Hence the coefficient of $D(\beta)$ in the analysis of $\Delta(\lambda)$ is $c_{(\alpha)}{}^{(\lambda)} \chi_{(\beta)}{}^{(\alpha)}$; in other words it is obtained by taking the indicated linear combination of those columns of the character table which correspond to classes (j) for which s^j appears in the product $q_{\lambda_1}(s) \cdots q_{\lambda_k}(s)$. This method is particularly suited to those partitions (λ) for which many of the λ_j are unity. For instance if they are all unity $q_{\lambda_1}(s) \cdots q_{\lambda_k}(s) = \{q_1(s)\}^n = s_1{}^n$ and so the coefficients in the analysis of $\Delta(1^n)$ are merely the characters of the unit class; in other words the coefficient of $D(\lambda)$ in the analysis of $\Delta(1^n)$ is the dimension of $D(\lambda)$; a fact of which we have independent knowledge since $\Delta(1^n)$ is the regular representation, of dimension $n!$, of the symmetric group. For $\Delta(2, 1^{n-2})$ the product $q_{\lambda_1}(s) \cdots q_{\lambda_k}(s)$ is $(s_1{}^n + s_1{}^{n-2} s_2) \div 2$ and hence the coefficient of $D(\lambda)$ in the analysis of $\Delta(2, 1^{n-2})$ is the mean of the characters $\chi_{(\lambda)}{}^{1^n}$, $\chi_{(\lambda)}{}^{(1^{n-2},2)}$ of the unit class and the transposition class, respectively. But it is clear that whilst the analysis of $\Delta(p, 1^{n-p})$ is relatively simple by this method a partition with three or more parts ≥ 2 leads to somewhat complicated calculations. Thus to make the analysis of $\Delta(3, 2^3)$ we would have to evaluate

$$q_3 q_2{}^3 = (s_1{}^3 + 3s_1 s_2 + 2s_3)(s_1{}^2 + s_2) \div 3!(2!)^3$$

and then form the indicated combination of the many columns of the character table (of the symmetric group on 9 letters) involved. We give below the

analysis of all $\Delta(\lambda)$ for $n \leqq 9$ and in making the calculations found the following method the most convenient. It rests on a knowledge of the analysis of the product $\{\lambda\}\{\mu\}$ of two simple characteristics which is given (for $\Sigma\lambda + \Sigma\mu \leqq 9$) in the following section. Suppose we wish to analyse the reducible representation $\Delta(3, 2, 1)$ of the symmetric group on six letters. We have $q_2q_1 = \{3\} + \{2, 1\}$ so that, since $q_3 = \{3\}$, $q_3q_2q_1 = \{3\}\{3\} + \{3\}\{2, 1\}$. From the values given at the end of the next section we read off

$$\begin{aligned}\{3\}\{3\} &= \{6\} + \{5,1\} + \{4,2\} + \{3^2\}.\\ \{3\}\{2,1\} &= \qquad\quad \{5,1\} + \{4,2\} + \{4,1^2\} + \{3,2,1\}\end{aligned}$$

so that

$$q_3q_2q_1 = \{6\} + 2\{5,1\} + 2\{4,2\} + \{4,1^2\} + \{3^2\} + \{3,2,1\}$$

or, equivalently,

$$\Delta(3,2,1) = D(6) + 2D(5,1) + 2D(4,2) + D(4,1^2) + D(3^2) + D(3,2,1).$$

In the following tables the irreducible representations are written across the top, the D being omitted in the interest of space, and the reducible representations are written down the left. For convenience of printing, Table 8, $n = 9$, is turned around so that the bottom of the page is the left-hand side of the table and the left-hand side of the page the top of the table. As examples of how the tables are read we cite the following:

$$\begin{aligned}
&n = 2;\ \Delta(1^2) = D(2) + D(1^2)\\
&n = 3;\ \Delta(2,1) = D(3) + D(2,1)\\
&n = 4;\ \Delta(2,1) = D(4) + 2D(3,1) + D(2^2) + D(2,1)\\
&n = 5;\ \Delta(2,1) = D(5) + 2D(4,1) + 2D(3,2) + D(3,1^2) + D(2^2,1)\\
&n = 6;\ \Delta(2^3) = D(6) + 2D(5,1) + 3D(4,2)\\
&\qquad\qquad + D(4,1^2) + D(3^2) + 2D(3,2,1) + D(2^3)
\end{aligned}$$

The numbers to the right of the main diagonal are all zero and are not written in. It may be observed that there is considerable duplication in the tables, the coefficients of the analysis of $\Delta(\lambda)$ being independent of n for the earlier partitions. Thus the coefficients in the analysis of the first twelve partitions of 9, from (9) to $(5, 1^4)$ inclusive, are the same as those in the analyses of the first twelve partitions of 8, from (8) to $(4, 1^4)$ inclusive. The table for $n = 10$ coincides with that for $n = 9$ for the first 19 partitions (from (10) to $(5, 1^5)$ inclusive) it being understood that the column under (5^2) is filled with 1's (from the partition (5^2) to $(5, 1^5)$); the correspondent to this column, namely $(4, 5)$, being absent from the table for $n = 9$. In completing the table for $n = 10$ it is convenient, in dealing with a four or five element partition which

ends in a 1 or 2, to use the corresponding partition in the table for $n = 9$ or 8, respectively. E. g. to obtain the analysis of $\Delta(4, 3, 2, 1)$ we use, from the table for $n = 9$, the result

$$q_4 q_3 q_2 = \{9\} + 2\{8, 1\} + 3\{7, 2\} + \{7, 1^2\} + 3\{6, 3\} + 2\{6, 2, 1\} \\ + 2\{5, 4\} + 2\{5, 3, 1\} + \{5, 2^2\} + \{4^2, 1\} + \{4, 3, 2\}.$$

Hence

$$q_4 q_3 q_2 q_1 = \{9\}\{1\} + 2\{8, 1\}\{1\} \cdots$$

and from the theorem concerning the analysis of the direct product of irreducible representations, given in the following section, we have

$$\{9\}\{1\} = \{10\} + \{9, 1\}; \qquad \{8, 1\}\{1\} = \{9, 1\} + \{8, 2\} + \{8, 1^2\} \text{ etc.,}$$

so that on collecting we obtain

$$\Delta(4, 3, 2, 1) = D(10) + 3D(9, 1) + 5D(8, 2) + 3D(8, 1^2) + 6D(7, 3) \\ + 6D(7, 2, 1) + D(7, 1^3) + 5D(6, 4) + 7D(6, 3, 1) \\ + 3D(6, 2^2) + 2D(6, 2, 1^2) + 2D(5^2) + 5D(5, 4, 1) \\ + 4D(5, 3, 2) + 2D(5, 3, 1^2) + D(5, 2^2, 1) + 2D(4^2, 2) \\ + D(4^2, 1^2) + D(4, 3^2) + D(4, 3, 2, 1).$$

Tables furnishing the analysis of $\Delta(\lambda)$ for values of n from 2 to 9 inclusive.

1. n = 2.

	(2)	(1^2)
Δ(2)	1	
Δ(1^2)	1	1

2. n = 3.

	(3)	(2,1)	(1^3)
Δ(3)	1		
Δ(2,1)	1	1	
Δ(1^3)	1	2	1

3. n = 4.

	(4)	(3,1)	(2^2)	(2,1^2)	(1^4)
Δ(4)	1				
Δ(3,1)	1	1			
Δ(2^2)	1	1	1		
Δ(2,1^2)	1	2	1	1	
Δ(1^4)	1	3	2	3	1

4. n = 5.

	(5)	(4,1)	(3,2)	(3,1^2)	(2^2,1)	(2,1^3)	(1^5)
Δ(5)	1						
Δ(4,1)	1	1					
Δ(3,2)	1	1	1				
Δ(3,1^2)	1	2	1	1			
Δ(2^2,1)	1	2	2	1	1		
Δ(2,1^3)	1	3	3	3	2	1	
Δ(1^5)	1	4	5	6	5	4	1

5. n = 6.

	(6)	(5,1)	(4,2)	(4,1^2)	(3^2)	(3,2,1)	(3,1^3)	(2^3)	(2^2,1^2)	(2,1^4)	(1^6)
Δ(6)	1										
Δ(5,1)	1	1									
Δ(4,2)	1	1	1								
Δ(4,1^2)	1	2	1	1							
Δ(3^2)	1	1	1	0	1						
Δ(3,2,1)	1	2	2	1	1	1					
Δ(3,1^3)	1	3	3	3	1	2	1				
Δ(2^3)	1	2	3	1	1	2	0	1			
Δ(2^2,1^2)	1	3	4	3	2	4	1	1	1		
Δ(2,1^4)	1	4	6	6	3	8	4	2	3	1	
Δ(1^6)	1	5	9	10	5	16	10	5	9	5	1

6. n=7.

	(7)	(6,1)	(5,2)	$(5,1^2)$	(4,3)	(4,2,1)	$(4,1^3)$	$(3^2,1)$	$(3,2^2)$	$(3,2,1^2)$	$(3,1^4)$	$(2^3,1)$	$(2^2,1^3)$	$(2,1^5)$	(1^7)
(7)	1														
$\Delta(6,1)$	1	1													
$\Delta(5,2)$	1	1	1												
$\Delta(5,1^2)$	1	2	1	1											
$\Delta(4,3)$	1	1	1	0	1										
$\Delta(4,2,1)$	1	2	2	1	1	1									
$\Delta(4,1^3)$	1	3	3	3	1	2	1								
$\Delta(3^2,1)$	1	2	2	1	2	1	0	1							
$\Delta(3,2^2)$	1	2	3	1	2	2	0	1	1						
$\Delta(3,2,1^2)$	1	3	4	3	3	4	1	2	1	1					
$\Delta(3,1^4)$	1	4	6	6	4	8	4	3	2	3	1				
$\Delta(2^3,1)$	1	3	5	3	4	6	1	3	3	2	0	1			
$\Delta(2^2,1^3)$	1	4	7	6	6	11	4	6	5	6	1	2	1		
$\Delta(2,1^5)$	1	5	10	10	9	20	10	11	10	15	5	5	4	1	
$\Delta(1^7)$	1	6	14	15	14	35	20	21	21	35	15	14	14	6	1

7. n=8.

	(8)	(7,1)	(6,2)	$(6,1^2)$	(5,3)	(5,2,1)	$(5,1^3)$	(4^2)	(4,3,1)	$(4,2^2)$	$(4,2,1^2)$	$(4,1^4)$	$(3^2,2)$	$(3^2,1^2)$	$(3,2^2,1)$	$(3,2,1^3)$	$(3,1^5)$	(2^4)	$(2^3,1^2)$	$(2^2,1^4)$	$(2,1^6)$	(1^8)
$\Delta(8)$	1																					
$\Delta(7,1)$	1	1																				
$\Delta(6,2)$	1	1	1																			
$\Delta(6,1^2)$	1	2	1	1																		
$\Delta(5,3)$	1	1	1	0	1																	
$\Delta(5,2,1)$	1	2	2	1	1	1																
$\Delta(5,1^3)$	1	3	3	3	1	2	1															
$\Delta(4^2)$	1	1	1	0	1	0	0	1														
$\Delta(4,3,1)$	1	2	2	1	2	1	0	1	1													
$\Delta(4,2^2)$	1	2	3	1	2	2	0	1	1	1												
$\Delta(4,2,1^2)$	1	3	4	3	3	4	1	1	2	1	1											
$\Delta(4,1^4)$	1	4	6	6	4	8	4	1	3	2	3	1										
$\Delta(3^2,2)$	1	2	3	1	3	2	0	1	2	1	0	0	1									
$\Delta(3^2,1^2)$	1	3	4	3	4	4	1	2	4	1	1	0	1	1								
$\Delta(3,2^2,1)$	1	3	5	3	5	6	1	2	5	3	2	0	2	1	1							
$\Delta(3,2,1^3)$	1	4	7	6	7	11	4	3	9	5	6	1	3	3	2	1						
$\Delta(3,1^5)$	1	5	10	10	10	20	10	4	15	10	15	5	5	6	5	4	1					
$\Delta(2^4)$	1	3	6	3	6	8	1	3	7	6	3	0	3	2	3	0	0	1				
$\Delta(2^3,1^2)$	1	4	8	6	9	14	4	4	13	9	9	1	6	5	6	2	0	1	1			
$\Delta(2^2,1^4)$	1	5	11	10	13	24	10	6	23	16	21	5	11	12	13	8	1	2	3	1		
$\Delta(2,1^6)$	1	6	15	15	19	40	20	9	40	30	45	15	21	26	30	24	6	5	9	5	1	
$\Delta(1^8)$	1	7	20	21	28	64	35	14	70	56	90	35	42	56	70	64	21	14	28	20	7	1

8. $n=9$.

	(9)	(8,1)	(7,2)	$(7,1^2)$	(6,3)	(6,2,1)	$(6,1^3)$	(5,4)	(5,3,1)	$(5,2^2)$	$(5,2,1^2)$	$(5,1^4)$	$(4^2,1)$	(4,3,2)	$(4,3,1^2)$	$(4,2^2,1)$	$(4,2,1^3)$	$(4,1^5)$	(3^3)	$(3^2,2,1)$	$(3^2,1^3)$	$(3,2^3)$	$(3,2^2,1^2)$	$(3,2,1^4)$	$(3,1^6)$	$(2^4,1)$	$(2^3,1^3)$	$(2^2,1^5)$	$(2,1^7)$	(1^9)
$\Delta(9)$	1																													
$\Delta(8,1)$	1	1																												
$\Delta(7,2)$	1	1	1																											
$\Delta(7,1^2)$	1	2	1	1																										
$\Delta(6,3)$	1	1	1	0	1																									
$\Delta(6,2,1)$	1	2	2	1	1	1																								
$\Delta(6,1^3)$	1	3	3	3	1	2	1																							
$\Delta(5,4)$	1	1	1	0	1	0	0	1																						
$\Delta(5,3,1)$	1	2	2	1	2	1	0	1	1																					
$\Delta(5,2^2)$	1	2	3	1	2	2	0	1	1	1																				
$\Delta(5,2,1^2)$	1	3	4	3	3	4	1	1	2	1	1																			
$\Delta(5,1^4)$	1	4	6	6	4	8	4	1	3	2	3	1																		
$\Delta(4^2,1)$	1	2	2	1	2	1	0	2	1	0	0	0	1																	
$\Delta(4,3,2)$	1	2	3	1	3	2	0	2	2	1	0	0	1	1																
$\Delta(4,3,1^2)$	1	3	4	3	4	4	1	3	4	1	1	0	2	1	1															
$\Delta(4,2^2,1)$	1	3	5	3	5	6	1	3	5	3	2	0	2	2	1	1														
$\Delta(4,2,1^3)$	1	4	7	6	7	11	4	4	9	5	6	1	3	3	3	2	1													
$\Delta(4,1^5)$	1	5	10	10	10	20	10	5	15	10	15	5	4	5	6	5	4	1												
$\Delta(3^3)$	1	2	3	1	4	2	0	2	3	1	0	0	1	2	0	0	0	0	1											
$\Delta(3^2,2,1)$	1	3	5	3	6	6	1	4	7	3	2	0	3	4	2	1	0	0	1	1										
$\Delta(3^2,1^3)$	1	4	7	6	8	11	4	6	12	5	6	1	6	6	6	2	1	0	1	2	1									
$\Delta(3,2^3)$	1	3	6	3	7	8	1	5	9	6	3	0	4	6	3	3	0	0	1	2	0	1								
$\Delta(3,2^2,1^2)$	1	4	8	6	10	14	4	7	16	9	9	1	7	10	8	6	2	0	2	4	1	1	1							
$\Delta(3,2,1^4)$	1	5	11	10	14	24	10	10	27	16	21	5	12	17	18	13	8	1	3	8	4	2	3	1						
$\Delta(3,1^6)$	1	6	15	15	20	40	20	14	45	30	45	15	19	30	36	30	24	6	5	16	10	5	9	5	1					
$\Delta(2^4,1)$	1	4	9	6	12	17	4	9	21	14	12	1	10	16	12	12	3	0	3	8	2	4	3	0	0	1				
$\Delta(2^3,1^3)$	1	5	12	10	17	28	10	13	36	23	27	5	17	28	27	24	12	1	6	17	7	7	9	2	0	2	1			
$\Delta(2^2,1^5)$	1	6	16	15	24	45	20	19	60	40	55	15	29	50	56	50	34	6	11	36	20	15	24	10	1	5	4	1		
$\Delta(2,1^7)$	1	7	21	21	34	70	35	28	99	70	105	35	49	91	111	105	84	21	21	77	50	35	63	35	7	14	14	6	1	
$\Delta(1^9)$	1	8	27	28	48	105	56	42	162	120	189	70	84	168	216	216	189	56	42	168	120	84	162	105	28	42	48	27	8	1

The direct product of irreducible representations. We consider two sets of n and m letters, neither set having a common letter so that the number of distinct letters in the two sets taken together is $n+m$. If (λ) is an arbitrary partition of n and (μ) an arbitrary partition of m and $D(\lambda)$, $D(\mu)$ the attached irreducible representations of the symmetric groups on n and m letters, respectively, the direct product $D(\lambda)\cdot D(\mu)$ is a representation, in general reducible, of the symmetric group on $n+m$ letters whose characteristic is the product $\{\lambda\}\{\mu\}$ of the characteristics of $D(\lambda)$ and $D(\mu)$. Our problem is the analysis of $D(\lambda)\cdot D(\mu)$ into its irreducible components. For the sake of brevity we shall omit the symbols D and write $(\lambda)\cdot(\mu)$ for $D(\lambda)\cdot D(\mu)$. If (ν) is a typical partition of $n+m$ a relation $(\lambda)\cdot(\mu)=\sum_{(\nu)} c_{(\nu)}(\nu)$ implies $\{\lambda\}\{\mu\}=\sum_{(\nu)} c_{(\nu)}\{\nu\}$ since $\{\lambda\}\{\mu\}$ is the characteristic of $(\lambda)\cdot(\mu)$. In order to arrive at a solution of our problem we first remark that the fundamental recurrence formula (13) may be generalised as follows. Let ξ_j be an operator which decreases the j-th member λ_j of the partition

$$(\lambda)=(\lambda_1,\cdots,\lambda_k)=(\lambda_1,\cdots,\lambda_n)$$

by unity $(j=1,\cdots,n)$. Then

$$\xi_j{}^p\{\lambda_1,\cdots,\lambda_k\}=\xi_j{}^p\{\lambda_1,\cdots,\lambda_n\}=\{\lambda_1,\cdots,\lambda_j-p,\cdots,\lambda_n\}$$

so that $\xi_j{}^p\{\lambda_1,\cdots,\lambda_k\}=0$ if $j>k$ for then $\{\lambda_1,\cdots,\lambda_j-p,\cdots,\lambda_n\}$ ends in a negative integer after the zeros at the end have been discarded. On writing

$S_p=\xi_1{}^p+\cdots+\xi_n{}^p$ we have

$$S_p\{\lambda_1,\cdots,\lambda_k\}=\{\lambda_1-p,\cdots,\lambda_k\}+\cdots+\{\lambda_1,\cdots,\lambda_k-p\}$$

so that our formula (13) may be written in the form

$$\{\lambda_1,\cdots,\lambda_k\}_{(\alpha)}=S_p\{\lambda_1,\cdots,\lambda_k\}_{(\alpha')}$$

and we may say that we have stripped off one cycle of p letters from (α). Following this by stripping from (α') a cycle of q letters we obtain

$$\{\lambda_1,\cdots,\lambda_k\}_{\alpha}=S_qS_p\{\lambda_1,\cdots,\lambda_k\}_{(\alpha'')}$$

where (α'') is the class, of the symmetric group on $n-p-q$ letters, which contains one less cycle on p letters and one less cycle on q letters than the class (α) of the symmetric group on n letters. More generally we may strip

off β_1 unary cycles, β_2 binary cycles etc.; to write the corresponding generalisation of the recurrence relation (13) it is a little more convenient to change the notation slightly so that n is replaced by $n+m$. Then (α) is a class of $n+m$ and we strip off β_1 unary cycles, β_2 binary cycles, $\cdots \beta_n$ n-ary cycles where (β) is a class of n. Denoting by (γ) the class of m which is such that $(\beta)+(\gamma)=(\alpha)$ our generalised recurrence relation appears in the form

$$\{\lambda\}_{(\alpha)} = S_1^{\beta_1} \cdots S_n^{\beta_n}\{\lambda\}_{(\gamma)}.$$

By means of (3) this may be written in the form

$$\{\lambda\}_{(\alpha)} = \sum_{(\theta)} \chi_{(\theta)}{}^{(\beta)} \phi_\theta(S)\{\lambda\}_{(\gamma)} \tag{19}$$

where the summation is over the partitions (θ) of n and (β), (γ) are any classes of n and m, respectively, whose sum is the class (α) of $n+m$.

Let now

$$\phi_{(\epsilon)}(s) = \frac{1}{n!}\sum_{(\delta)} N_{(\delta)}\chi_{(\epsilon)}{}^{(\delta)} s^{(\delta)}$$

be any simple characteristic of the symmetric group on n letters (so that (ϵ) is a partition of n) and, similarly, let

$$\phi_{(\nu)}(s) = \frac{1}{m!}\sum_{(\tau)} M_{(\tau)}\chi_{(\nu)}{}^{(\tau)} s^{(\tau)}$$

be any simple characteristic of the symmetric group on m letters, (ν) being a partition of m; their product is

$$\phi_{(\epsilon)}(s)\phi_{(\nu)}(s) = \frac{1}{n!}\,\frac{1}{m!}\sum_{(\delta)(\tau)} N_{(\delta)}M_{(\tau)}\chi_{(\epsilon)}{}^{(\delta)}\chi_{(\nu)}{}^{(\tau)} s^{(\delta)+(\tau)}$$

and since $(\delta)+(\tau)$ is a partition of $n+m$ we have, from (3),

$$s^{(\delta)+(\tau)} = \sum_{(\alpha)} \chi_{(\alpha)}{}^{(\delta)+(\tau)} \phi_{(\alpha)}(s)$$

the summation being over all partitions (α) of $n+m$. On substituting for $\chi_{(\alpha)}{}^{(\delta)+(\tau)}$ its value $\sum_{\theta} \chi_{(\theta)}{}^{(\delta)}\{\phi_\theta(S)\boldsymbol{\chi}_\alpha\}^{(\tau)}$ from (19) and summing with respect to (δ) we obtain, in view of the orthogonality relations (1) between the characters of the symmetric group on n letters,

$$\phi_{(\epsilon)}(s)\phi_{(\nu)}(s) = \frac{1}{m!}\sum_{(\alpha)(\tau)} M_{(\tau)}\chi_{(\nu)}{}^{(\tau)}\{\phi_{(\epsilon)}(S)\boldsymbol{\chi}_{(\alpha)}\}^{(\tau)}\phi_{(\alpha)}(s) \cdots. \tag{20}$$

Now $\phi_{(\epsilon)}(S)$ is a symmetric function, of degree n, with integral coefficients, of the $n+m$ operators ξ_j and so is of the form $\sum_{(\pi)} c_\pi[\xi_1^{\pi_1}\cdots\xi_n^{\pi_n}]$ where (π) is a partition of n and $[\xi_1^{\pi_1}\xi_2^{\pi_2}\cdots\xi_n^{\pi_n}]$ denotes the symmetric function of the $n+m$ operators $(\xi_1,\cdots,\xi_{n+m})$ whose leading term is $\xi_1^{\pi_1}\xi_2^{\pi_2}\cdots\xi_n^{\pi_n}$. The result of operating with $\xi_1^{\pi_1}\cdots\xi_n^{\pi_n}$ on $\chi_{(a)}$ is

$$\chi_{(a_1-\pi_1,\,a_2-\pi_2,\,\ldots,\,a_n-\pi_n,\,a_{n+1},\,\ldots,\,a_{n+m})}$$

and we may denote the result of operating with $[\xi_1^{\pi_1}\cdots\xi_n^{\pi_n}]$ on $\chi_{(a)}$ by $\chi_{[(a)-(\pi)]}$. The summation with respect to τ yields zero, owing to the orthogonality relations between the characters of the symmetric group on m letters, save when (α) is such that one member of $[(\alpha)-(\pi)]$ is the same as (ν), with the same convention as before regarding the rearrangement of disordered partitions; in which case the coefficient of $\phi_{(a)}(s)$ is $c_{(\pi)}$. The simplest examples may serve to make the theory clear; thus let $(\epsilon)=(1)$ so that we wish to analyse $\{1\}\{\nu_1,\cdots,\nu_k\}$; the only $\phi_{(a)}(s)$ which appear in this product are those for which (α) is obtained from $(\nu_1,\cdots,\nu_k,0)$ by adding unity to one of its members and these all occur with coefficient unity; for

$$\phi_{(1)}(S)=S_1=\xi_1+\xi_2+\cdots\xi_{n+1}.$$

Hence

$$\{1\}\{\nu_1,\cdots,\nu_k\}=\{\nu_1+1,\nu_2,\cdots,\nu_k\}+\cdots+\{\nu_1,\cdots,\nu_k+1\}+\{\nu_1,\cdots,\nu_k,1\}.$$

E. g.,
$$\begin{aligned}\{1\}\{4,2^2\}&=\{5,2^2\}+\{4,3,2\}+\{4,2,3\}+\{4,2^2,1\}\\&=\{5,2^2\}+\{4,3,2\}+\{4,2^2,1\}\end{aligned}$$

or, equivalently $(1)\cdot(4,2^2)=(5,2^2)+(4,3,2)+(4,2^2,1)$. The next simplest example is furnished by $\{2\}\{\nu_1,\cdots,\nu_k\}$; here

$$\phi_2(S)=p_2(\xi)=\Sigma\xi_j^2+\Sigma\xi_j\xi_p$$

and so

$$\begin{aligned}\{2\}\{\nu_1,\cdots,\nu_k\}=\{\nu_1+2,\cdots,\nu_k\}&+\cdots+\{\nu_1,\cdots,\nu_k,2\}\\&+\{\nu_1+1,\nu_2+1,\cdots,\nu_k\}+\cdots\\&+\{\nu_1+1,\cdots,\nu_k,1\}+\cdots\end{aligned}$$

the terms $\{\nu_1+1,\cdots,\nu_k,0,1\}$ vanishing and the terms $\{\nu_1,\cdots,\nu_k,0,2\}$ and $\{\nu_1,\cdots,\nu_k,1,1\}$ cancelling one another.

E. g.,
$$\begin{aligned}\{2\}\{3,2\}&=\{5,2\}+\{3,4\}+\{3,2^2\}+\{4,3\}+\{4,2,1\}+\{3^2,1\}\\&=\{5,2\}+\{4,3\}+\{4,2,1\}+\{3^2,1\}+\{3,2^2\}.\end{aligned}$$

Similarly since

$$\phi_3(S) = p_3(\xi) = \Sigma\xi_j^3 + \Sigma\xi_j^2\xi_l + \Sigma\xi_j\xi_l\xi_m$$

we have

$$\begin{aligned}\{3\}\{\nu_1, \cdots, \nu_k\} = \{\nu_1 + 3, \cdots, \nu_k\} &+ \cdots + \{\nu_1, \cdots, \nu_k, 3\} \\ &+ \{\nu_1 + 2, \nu_2 + 1, \cdots, \nu_k\} + \cdots \\ &+ \{\nu_1 + 2, \cdots, \nu_k, 1\} + \cdots \\ &+ \{\nu_1 + 1, \cdots, \nu_k, 2\} + \cdots \\ &+ \{\nu_1 + 1, \nu_2 + 1, \nu_3 + 1, \cdots, \nu_k\} + \cdots \\ &+ \{\nu_1 + 1, \nu_2 + 1, \cdots, \nu_k, 1\} + \cdots .\end{aligned}$$

(the remaining terms vanishing or cancelling each other).

E. g., $$\begin{aligned}\{3\}\{2, 1^2\} &= \{5, 1^2\} + \{2, 4, 1\} + \{2, 1, 4\} + \{2, 1^2, 3\} \\ &+ \{4, 2, 1\} + \{3^2, 1\} + \{4, 1, 2\} \\ &\qquad + \{3, 1, 3\} + \{2, 3, 2\} + \{2^2, 3\} \\ &+ \{4, 1^3\} + \{2, 3, 1^2\} + \{2, 1, 3, 1\} \\ &\qquad + \{3, 1^2, 2\} + \{2^2, 1, 2\} + \{2, 1, 2^2\} \\ &+ \{3, 2^2\} + \{3, 2, 1^2\} + \{3, 1, 2, 1\} + \{2^3, 1\} \\ &= \{5, 1^2\} + \{4, 2, 1\} + \{4, 1^3\} + \{3, 2, 1^2\}.\end{aligned}$$

Since $\phi_{(1^2)}(S) = \sigma_2(\xi) = \Sigma\xi_j\xi_l$ we have

$$\begin{aligned}\{1^2\}\{\nu, \cdots, \nu_k\} &= \{\nu_1 + 1, \nu_2 + 1, \cdots, \nu_k\} \\ &\quad + \{\nu_1 + 1, \cdots, \nu_k, 1\} + \cdots + \{\nu_1, \cdots, \nu_k, 1, 1\}\end{aligned}$$

E. g., $$\begin{aligned}\{1^2\}\{3, 2, 1\} &= \{4, 3, 1\} + \{4, 2^2\} + \{3^2, 2\} \\ &\quad + \{4, 2, 1^2\} + \{3^2, 1^2\} + \{3, 2^2, 1\} + \{3, 2, 1^3\}.\end{aligned}$$

It is clear that whilst this method is entirely practicable when one of the factors is $\{2\}$, $\{3\}$, $\{1^2\}$, $\{1^3\}$ it rapidly becomes very tedious in other cases. We give below tables of all direct products $(\lambda) \cdot (\mu)$ for which $n + m \leq 9$ and we found the following method, which is sufficiently illustrated by an example, entirely convenient. Suppose we wish $\{3, 1\}\{2^2\}$; we write $\{3, 1\} = \{3\}\{1\} - \{4\}$ (since $\begin{vmatrix} q_3 & q_4 \\ q_0 & q_1 \end{vmatrix} = q_3q_1 - q_4$) and we see that the calculation rests on that of $\{4\}\{2^2\}$. But $\{4, 2^2\} = \{4\}\{2^2\} - \{1\}\{5, 2\} + \{5, 3\}$ since

$$\begin{vmatrix} q_4 & q_5 & q_6 \\ q_1 & q_2 & q_3 \\ q_0 & q_1 & q_2 \end{vmatrix} = q_4 \begin{vmatrix} q_2 & q_3 \\ q_1 & q_2 \end{vmatrix} - q_1 \begin{vmatrix} q_5 & q_6 \\ q_1 & q_2 \end{vmatrix} + \begin{vmatrix} q_5 & q_6 \\ q_2 & q_3 \end{vmatrix} .$$

Hence

$$\{4\}\{2^2\} = \{4, 2^2\} + \{1\}\{5, 2\} - \{5, 3\}$$
$$= \{6, 2\} + \{5, 2, 1\} + \{4, 2^2\}.$$

Similarly

$$\{3\}\{2^2\} = \{5, 2\} + \{4, 2, 1\} + \{3, 2^2\}$$

and so

$$\begin{aligned}\{1\}\{3\}\{2\}^2 = {} & [\{6, 2\} + \{5, 3\} + \{5, 2, 1\}] \\ & + [\{5, 2, 1\} + \{4, 3, 1\} + \{4, 2^2\} + \{4, 2, 1^2\}] \\ & + [\{4, 2^2\} + \{3^2, 2\} + \{3, 2, 3\} + \{3, 2^2, 1\}] \\ = {} & \{6, 2\} + \{5, 3\} + 2\{5, 2, 1\} + \{4, 3, 1\} + 2\{4, 2^2\} + \{4, 2, 1^2\} \\ & + \{3^2, 2\} + \{3, 2^2, 1\}.\end{aligned}$$

Hence

$$\begin{aligned}\{3, 1\}\{2^2\} = {} & \{5, 3\} + \{5, 2, 1\} \\ & + \{4, 3, 1\} + \{4, 2^2\} + \{4, 2, 1^2\} + \{3^2, 2\} + \{3, 2^2, 1\}.\end{aligned}$$

In the following tables the irreducible representations are written across the top and the desired direct products are indicated down the left. As examples of how the tables are read we cite the following:

$$\begin{aligned} & n + m = 3;\ (2) \cdot (1) = (3) + (2, 1) \\ & n + m = 4;\ (2, 1) \cdot (1) = (3, 1) + (2^2) + (2, 1^2) \\ & n + m = 5;\ (3) \cdot (1^2) = (4, 1) + (3, 1^2) \\ & n + m = 6;\ (2, 1) \cdot (2, 1) = (4, 2) + (4, 1^2) + (3^2) \\ & \qquad\qquad + 2(3, 2, 1) + (2^3) + (3, 1^3) + (2^2, 1^2).\end{aligned}$$

Since a change of sign of $(s_2, s_4, \cdots)$ sends $\{\lambda\}$ into the associated characteristic $\{\mu\}$ it is clear that $\{\mu\}\{\mu'\}$ is obtained from $\{\lambda\}\{\lambda'\}$ by merely taking the associated characteristics or representations; e.g. from $\{3\}\{1^2\} = \{4, 1\} + \{3, 1^2\}$ we read $\{1^3\}\{2\} = \{2, 1^3\} + \{3, 1^2\}$. We use this trivially evident fact to materially cut down the size of the tables (without causing trouble to the user) by writing $\{\mu\}\{\mu'\}$ on the right side of the table directly opposite $\{\lambda\}\{\lambda'\}$—where $\{\mu\}$ and $\{\lambda\}$ are associated simple characteristics of the symmetric group on n letters whilst $\{\mu'\}$ and $\{\lambda'\}$ are associated simple characteristics of the symmetric group on m letters. It being understood that when we pick up our direct product on the right-hand side of the table we find the irreducible representations of the symmetric group on $n + m$ letters which occur in the

analysis of the direct product at the *bottom* of the table; whilst when we pick up the direct product on the left-hand side of the table we find the representations which occur in its analysis at the *top* of the table. For convenience of printing, the tables for $n + m = 8, 9$ have been turned so that the top of the page is the right-hand side of the table and the left-hand side of the page is the top of the table.

Tables furnishing the analysis of the direct product $(\lambda) \cdot (\lambda')$ for all values of $n + m$ from 2 to 9 inclusive.

1. n+m=2.

$$(1).(1) = (2) + (1^2)$$

2. n+m=3.

	(3)	$(2,1)$	(1^3)	
$(2).(1)$	1	1		$(1^2).(1)$
	(1^3)	$(2,1)$	(3)	

3. n+m=4.

	(4)	$(3,1)$	(2^2)	$(2,1^2)$	(1^4)	
$(3).(1)$	1	1				$(1^3).(1)$
$(2,1).(1)$		1	1	1		$(2,1).(1)$
$(2).(2)$	1	1	1			$(1^2).(1^2)$
$(2).(1^2)$		1		1		$(1^2).(2)$
	(1^4)	$(2,1^2)$	(2^2)	$(3,1)$	(4)	

4. n+m=5.

	(5)	$(4,1)$	$(3,2)$	$(3,1^2)$	$(2^2,1)$	$(2,1^3)$	(1^5)	
$(4).(1)$	1	1						$(1^4).(1)$
$(3,1).(1)$		1	1	1				$(2,1^2).(1)$
$(2^2).(1)$			1		1			$(2^2).(1)$
$(3).(2)$	1	1	1					$(1^3).(1^2)$
$(2,1).(2)$		1	1	1	1			$(2,1).(1^2)$
$(1^3).(2)$				1		1		$(3).(1^2)$
	(1^5)	$(2,1^3)$	$(2^2,1)$	$(3,1^2)$	$(3,2)$	$(4,1)$	(5)	

5. $n+m=6$.

	(6)	$(5,1)$	$(4,2)$	$(4,1^2)$	$(3)^2$	$(3,2,1)$	$(3,1^3)$	(2^3)	$(2^2,1^2)$	$(2,1^4)$	(1^6)	
$(5).(1)$	1	1										$(1^5).(1)$
$(4,1).(1)$		1	1	1								$(2,1^3).(1)$
$(3,2).(1)$			1		1	1						$(2^2,1).(1)$
$(3,1^2).(1)$				1		1	1					$(3,1^2).(1)$
$(4).(2)$	1	1	1									$(1^4).(1^2)$
$(3,1).(2)$		1	1	1	1	1						$(2,1^2).(1^2)$
$(2^2).(2)$			1			1		1				$(2^2).(1^2)$
$(2,1^2).(2)$				1		1	1		1			$(3,1).(1^2)$
$(1^4).(2)$							1			1		$(4).(1^2)$
$(3).(3)$	1	1	1		1							$(1^3).(1^3)$
$(3).(2,1)$		1	1	1		1						$(1^3).(2,1)$
$(3).(1^3)$				1			1					$(1^3).(3)$
$(2,1).(2,1)$			1	1	1	2	1	1	1			$(2,1).(2,1)$
	(1^6)	$(2,1^4)$	$(2^2,1^2)$	$(3,1^3)$	(2^3)	$(3,2,1)$	$(4,1^2)$	(3^2)	$(4,2)$	$(5,1)$	(6)	

6. $n+m=7$.

	(7)	$(6,1)$	$(5,2)$	$(5,1^2)$	$(4,3)$	$(4,2,1)$	$(4,1^3)$	$(3^2,1)$	$(3,2^2)$	$(3,2,1^2)$	$(3,1^4)$	$(2^3,1)$	$(2^2,1^3)$	$(2,1^5)$	(1^7)	
$(6).(1)$	1	1														$(1^6).(1)$
$(5,1).(1)$		1	1	1												$(2,1^4).(1)$
$(4,2).(1)$			1		1	1										$(2^2,1^2).(1)$
$(4,1^2).(1)$				1		1	1									$(3,1^3).(1)$
$(3^2).(1)$					1			1								$(2^3).(1)$
$(3,2,1).(1)$						1		1	1	1						$(3,2,1).(1)$
$(5).(2)$	1	1	1													$(1^5).(1^2)$
$(4,1).(2)$		1	1	1	1	1										$(2,1^3).(1^2)$
$(3,2).(2)$			1		1	1		1	1							$(2^2,1).(1^2)$
$(3,1^2).(2)$				1		1	1	1		1						$(3,1^2).(1^2)$
$(2^2,1).(2)$						1			1	1		1				$(3,2).(1^2)$
$(2,1^3).(2)$							1			1	1		1			$(4,1).(1^2)$
$(1^5).(2)$											1			1		$(5).(1^2)$
$(4).(3)$	1	1	1		1											$(1^4).(1^3)$
$(3,1).(3)$		1	1	1	1	1		1								$(2,1^2).(1^3)$
$(2^2).(3)$			1			1			1							$(2^2).(1^3)$
$(2,1^2).(3)$				1		1	1			1						$(3,1).(1^3)$
$(1^4).(3)$							1				1					$(4).(1^3)$
$(4).(2,1)$		1	1	1		1										$(1^4).(2,1)$
$(3,1).(2,1)$			1	1	1	2	1	1	1	1						$(2,1^2).(2,1)$
$(2^2).(2,1)$					1	1		1	1	1		1				$(2^2).(2,1)$
	(1^7)	$(2,1^5)$	$(2^2,1^3)$	$(3,1^4)$	$(2^3,1)$	$(3,2,1^2)$	$(4,1^3)$	$(3,2^2)$	$(3^2,1)$	$(4,2,1)$	$(5,1^2)$	$(4,3)$	$(5,2)$	$(6,1)$	(7)	

7. $n+m=8$

		$(1^7).(1)$ / $(7).(1)$	$(2,1^5).(1)$ / $(6,1).(1)$	$(2^2,1^3).(1)$ / $(5,2).(1)$	$(3,1^4).(1)$ / $(5,1^2).(1)$	$(2^3,1).(1)$ / $(4,3).(1)$	$(3,2,1^2).(1)$ / $(4,2,1).(1)$	$(4,1^3).(1)$ / $(4,1^3).(1)$	$(3,2^2).(1)$ / $(3^2,1).(1)$	$(1^6).(1^2)$ / $(6).(2)$	$(2,1^4).(1^2)$ / $(5,1).(2)$	$(2^2,1^2).(1^2)$ / $(4,2).(2)$	$(3,1^3).(1^2)$ / $(4,1^2).(2)$	$(2^3).(1^2)$ / $(3^2).(2)$	$(3,2,1).(1^2)$ / $(3,2,1).(2)$	$(4,1^2).(1^2)$ / $(3,1^3).(2)$	$(3^2).(1^2)$ / $(2^3).(2)$	$(4,2).(1^2)$ / $(2^2,1^2).(2)$	$(5,1).(1^2)$ / $(2,1^4).(2)$	$(6).(1^2)$ / $(1^6).(2)$	$(1^5).(1^3)$ / $(5).(3)$	$(2,1^3).(1^3)$ / $(4,1).(3)$	$(2^2,1).(1^3)$ / $(3,2).(3)$	$(3,1^2).(1^3)$ / $(3,1^2).(3)$	$(3,2).(1^3)$ / $(2^2,1).(3)$	$(4,1).(1^3)$ / $(2,1^3).(3)$	$(5).(1^3)$ / $(1^5).(3)$	$(1^5).(2,1)$ / $(5).(2,1)$	$(2,1^3).(2,1)$ / $(4,1).(2,1)$	$(2^2,1).(2,1)$ / $(3,2).(2,1)$	$(3,1^2).(2,1)$ / $(3,1^2).(2,1)$	$(1^4).(1^4)$ / $(4).(4)$	$(2,1^2).(1^4)$ / $(3,1).(4)$	$(2^2).(1^4)$ / $(2^2).(4)$	$(3,1).(1^4)$ / $(2,1^2).(4)$	$(4).(1^4)$ / $(1^4).(4)$	$(2,1^2).(2,1^2)$ / $(3,1).(3,1)$	$(2^2).(2,1^2)$ / $(2^2).(3,1)$	$(3,1).(2,1^2)$ / $(2,1^2).(3,1)$	$(2^2).(2^2)$ / $(2^2).(2^2)$
(8)	(1^8)																																							
$(7,1)$	$(2,1^6)$																			1																				
$(6,2)$	$(2^2,1^4)$																		1																					
$(5,3)$	$(2^3,1^2)$																	1																						
(4^2)	(2^4)																1																							1
$(6,1^2)$	$(3,1^5)$																		1	1							1													
$(5,2,1)$	$(3,2,1^3)$															1		1	1							1					1								1	
$(4,3,1)$	$(3,2^2,1)$														1		1	1							1					1	1							1	1	1
$(4,2^2)$	$(3^2,1^2)$								1						1	1								1						1	1						1		1	1
$(3^2,2)$	$(3^2,2)$								1					1	1								1							1	1						1	1		
$(5,1^3)$	$(4,1^4)$							1								1			1							1	1				1					1			1	
$(4,2,1^2)$	$(4,2,1^2)$						1	1					1		1	1		1						1	1	1			1	1	2				1		1	1	2	
$(3^2,1^2)$	$(4,2^2)$						1					1			1		1						1		1				1	1	1			1			1	1	1	1
$(3,2^2,1)$	$(4,3,1)$					1	1		1			1	1	1	1							1	1	1					1	2	1		1				2	1	1	1
(2^4)	(4^2)					1						1										1								1		1					1			1
$(4,1^4)$	$(5,1^3)$				1			1					1			1								1		1			1		1				1	1	1		1	
$(3,2,1^3)$	$(5,2,1)$			1	1		1				1	1	1		1							1	1	1	1			1	2	1	1		1	1	1		2	1	1	
$(2^3,1^2)$	$(5,3)$			1		1					1	1		1							1	1	1						1	1		1	1				1	1		
$(3,1^5)$	$(6,1^2)$		1		1						1		1									1		1				1	1				1		1		1			
$(2^2,1^4)$	$(6,2)$		1	1						1	1	1									1	1	1					1	1			1	1	1			1			
$(2,1^6)$	$(7,1)$	1	1							1	1										1	1						1				1	1							
(1^8)	(8)	1								1											1											1								

8. $n+m=9$.

	$(1)^8.(1)$ / $(8).(1)$	$(2,1^6).(1)$ / $(7,1).(1)$	$(2^2,1^4).(1)$ / $(6,2).(1)$	$(3,1^5).(1)$ / $(6,1^2).(1)$	$(2^3,1^2).(1)$ / $(5,3).(1)$	$(3,2,1^3).(1)$ / $(5,2,1).(1)$	$(4,1^4).(1)$ / $(5,1^3).(1)$	$(2^4).(1)$ / $(4^2).(1)$	$(3,2^2,1).(1)$ / $(4,3,1).(1)$	$(3^2,1^2).(1)$ / $(4,2^2).(1)$	$(4,2,1^2).(1)$ / $(4,2,1^2).(1)$	$(3^2,2).(1)$ / $(3^2,2).(1)$	$(1^7).(1^2)$ / $(7).(2)$	$(2,1^5).(1^2)$ / $(6,1).(2)$	$(2^2,1^3).(1^2)$ / $(5,2).(2)$	$(3,1^4).(1^2)$ / $(5,1^2).(2)$	$(2^3,1).(1^2)$ / $(4,3).(2)$	$(3,2,1^2).(1^2)$ / $(4,2,1).(2)$	$(4,1^3).(1^2)$ / $(4,1^3).(2)$	$(3,2^2).(1^2)$ / $(3^2,1).(2)$	$(3^2,1).(1^2)$ / $(3,2^2).(2)$	$(4,2,1).(1^2)$ / $(3,2,1^2).(2)$	$(5,1^2).(1^2)$ / $(3,1^4).(2)$	$(4,3).(1^2)$ / $(2^3,1).(2)$	$(5,2).(1^2)$ / $(2^2,1^3).(2)$	$(6,1).(1^2)$ / $(2,1^5).(2)$	$(7).(1^2)$ / $(1^7).(2)$	$(1^6).1^3$ / $(6).(3)$	$(2,1^4).(1^3)$ / $(5,1).(3)$	$(2^2,1^2).)1^3)$ / $(4,2).(3)$	$(3,1^3).(1^3)$ / $(4,1^2).(3)$	
(1^9)																																(9)
$(2,1^7)$																											1					$(8,1)$
$(2^2,1^5)$																										1						$(7,2)$
$(2^3,1^3)$																									1							$(6,3)$
$(2^4,1)$																								1								$(5,4)$
$(3,1^6)$																										1	1					$(7,1^2)$
$(3,2,1^4)$																							1		1	1						$(6,2,1)$
$(3,2^2,1^2)$																						1		1	1							$(5,3,1)$
$(3,2^3)$																					1			1								$(4^2,1)$
$(3^2,1^3)$																						1	1									$(5,2^2)$
$(3^2,2,1)$												1								1	1	1										$(4,3,2)$
(3^3)												1								1												(3^3)
$(4,1^5)$																							1			1						$(6,1^3)$
$(4,2,1^3)$											1								1			1	1		1							$(5,2,1^2)$
$(4,2^2,1)$										1	1							1			1	1		1								$(4,3,1^2)$
$(4,3,1^2)$									1		1							1	1	1		1									1	$(4,2^2,1)$
$(4,3,2)$									1	1		1					1	1		1	1									1		$(3^2,2,1)$
$(4^2,1)$								1	1								1	1												1	1	$(3,2^3)$
$(5,1^4)$							1												1				1									$(5,1^4)$
$(5,2,1^2)$						1	1				1					1		1	1			1									1	$(4,2,1^3)$
$(5,2^2)$						1				1					1			1			1									1		$(3^2,1^3)$
$(5,3,1)$					1	1			1						1	1	1	1		1									1	1	1	$(3,2^2,1^2)$
$(5,4)$					1			1							1		1												1	1		$(2^4,1)$
$(6,1^3)$				1			1									1			1												1	$(4,1^5)$
$(6,2,1)$			1	1		1								1	1	1		1											1	1	1	$(3,2,1^4)$
$(6,3)$			1		1									1	1		1											1	1	1		$(2^3,1^3)$
$(7,1^2)$		1		1										1		1													1		1	$(3,1^6)$
$(7,2)$		1	1										1	1	1													1	1	1		$(2^2,1^5)$
$(8,1)$	1	1											1	1														1	1			$(2,1^7)$
(9)	1												1															1				(1^9)

	$(2^3).(1^3)$	$(3,2,1).(1^3)$	$(4,1^2).(1^3)$	$(3^2).(1^3)$	$(4,2).(1^3)$	$(5,1).(1^3)$	$(6).(1^3)$	$(1^6).(2,1)$	$(2,1^4).(2,1)$	$(2^2,1^2).(2,1)$	$(3,1^3).(2,1)$	$(2^3).(2,1)$	$(3,2,1).(2,1)$	$(1^5).(1^4)$	$(2,1^3).(1^4)$	$(2^2,1).(1^4)$	$(3,1^2).(1^4)$	$(3,2).(1^4)$	$(4,1).(1^4)$	$(5).(1^4)$	$(1^5).(2,1^2)$	$(2,1^3).(2,1^2)$	$(2^2,1).(2,1^2)$	$(3,1^2).(2,1^2)$	$(3,2).(2,1^2)$	$(4,1).(2,1^2)$	$(5).(2,1^2)$	$(1^5).(2^2)$	$(2,1^3).(2^2)$	$(2^2,1).(2^2)$	$(3,1^2).(2^2)$	
(1^9)																																(9)
$(2,1^7)$																																$(8,1)$
$(2^2,1^5)$																																$(7,2)$
$(2^3,1^3)$																																$(6,3)$
$(2^4,1)$																																$(5,4)$
$(3,1^6)$							1																				1					$(7,1^2)$
$(3,2,1^4)$						1																				1	1					$(6,2,1)$
$(3,2^2,1^2)$					1								1												1	1					1	$(5,3,1)$
$(3,2^3)$				1									1												1					1		$(4^2,1)$
$(3^2,1^3)$			1										1											1		1						$(5,2^2)$
$(3^2,2,1)$		1										1	2										1	1	1					1	1	$(4,3,2)$
(3^3)	1												1										1								1	(3^3)
$(4,1^5)$						1	1													1						1	1					$(6,1^3)$
$(4,2,1^3)$			1		1	1					1		1						1					1	1	2	1				1	$(5,2,1^2)$
$(4,2^2,1)$		1		1	1					1	1		2					1					1	1	2	1			1	1	1	$(4,3,1^2)$
$(4,3,1^2)$		1	1							1	1	1	2				1					1	1	2	1	1				1	1	$(4,2^2,1)$
$(4,3,2)$	1	1								1	1	1	2			1						1	2	1	1				1	1	1	$(3^2,2,1)$
$(4^2,1)$										1		1	1		1							1	1	1						1		$(3,2^3)$
$(5,1^4)$			1			1					1								1	1				1		1						$(5,1^4)$
$(5,2,1^2)$		1	1		1				1	1	2		1				1	1	1			1	1	2	1	1			1		1	$(4,2,1^3)$
$(5,2^2)$		1		1					1	1	1		1			1		1				1	1	1	1			1	1	1		$(3^2,1^3)$
$(5,3,1)$	1	1							1	2	1	1	1		1	1	1				1	2	2	1	1				1	1	1	$(3,2^2,1^2)$
$(5,4)$										1		1		1	1							1	1							1		$(2^4,1)$
$(6,1^3)$			1						1		1						1		1			1		1								$(4,1^5)$
$(6,2,1)$		1						1	2	1	1				1	1	1	1			1	2	1	1				1	1			$(3,2,1^4)$
$(6,3)$	1								1	1				1	1	1					1	1	1						1			$(2^3,1^3)$
$(7,1^2)$								1	1						1		1				1	1										$(3,1^6)$
$(7,2)$								1	1					1	1	1					1	1						1				$(2^2,1^5)$
$(8,1)$								1						1	1						1											$(2,1^7)$
(9)														1																		(1^9)
	$(3^2).(3)$	$(3,2,1).(3)$	$(3,1^3).(3)$	$(2^3).(3)$	$(2^2,1^2).(3)$	$(2,1^4).(3)$	$(1^6).(3)$	$(6).(2,1)$	$(5,1).(2,1)$	$(4,2).(2,1)$	$(4,1^2).(2,1)$	$(3^2).(2,1)$	$(3,2,1).(2,1)$	$(5).(4)$	$(4,1).(4)$	$(3,2).(4)$	$(3,1^2).(4)$	$(2^2,1).(4)$	$(2,1^3).(4)$	$(1^5).(4)$	$(5).(3,1)$	$(4,1).(3,1)$	$(3,2).(3,1)$	$(3,1^2).(3,1)$	$(2^2,1).(3,1)$	$(2,1^3).(3,1)$	$(1^5).(3,1)$	$(5).(2^2)$	$(4,1).(2^2)$	$(3,2).(2^2)$	$(3,1^2).(2^2)$	

JOHNS HOPKINS UNIVERSITY,
INSTITUTE FOR ADVANCED STUDY.

REFERENCES.

1. G. Frobenius, "Über Gruppencharaktere," *Berliner Berichte* (1896), pp. 985-1021.
2. ———, "Über Relationen zwischen den Charakteren einer Gruppe und denen ihrer Untergruppen," *Berliner Berichte* (1898), pp. 501-515.
3. ———, "Über die Composition der Charaktere einer Gruppe," *Berliner Berichte* (1899), pp. 330-339.
4. ———, "Über die Charaktere der symmetrischen Gruppe," *Berliner Berichte* (1900), pp. 516-534.
5. ———, "Über die charakterischen Einheiten der symmetrischen Gruppe," *Berliner Berichte* (1903), pp. 328-358.
6. ———, "Über die Charaktere der mehrfach transitiven Gruppen," *Berliner Berichte* (1904), pp. 528-571.
7. I. Schur, "Über einer Klasse von Matrizen," *Dissertation*, Berlin (1901).
8. ———, "Neue Begründung der Theorie der Gruppencharaktere," *Berliner Berichte* (1905), pp. 406-432.
9. ———, "Über die Darstellung der symmetrischen Gruppe durch lineare homogene Substitutionen," *Berliner Berichte* (1908), pp. 664-678.
10. ———, "Über die rationalen Darstellungen der allgemeinen linearen Gruppe," *Berliner Berichte* (1927), pp. 58-75.
11. ———, "Die algebraischen Grundlagen der Darstellungstheorie der Gruppen," *Zurich Lectures* (1936), pp. 1-74.
12. A. Young, "On quantitative substitutional analysis," Part 1, *Proceedings of the London Mathematical Society* (1), **32** (1901), pp. 384-404; Part II, *ibid.* (1) **34** (1902), pp. 202-208; Part III, *ibid.* (2), **28** (1928), pp. 255-292; Parts IV, V, *ibid.* (2), **31** (1930), pp. 253-272, 273-288; Part VI, *ibid.* (2), **34** (1932), pp. 196-230; Part VII, *ibid.* (2), **36** (1934), pp. 304-368; Part VIII, *ibid.* (2), **37** (1934), pp. 441-495.
13. D. E. Littlewood and A. R. Richardson, "Group characters and algebra," *Philosophical Transactions of the Royal Society of London* (A), **233** (1934), pp. 99-141.
14. ———, "Immanants of some special matrices," *Quarterly Journal of Mathematics*, Oxford Ser. **5** (1934), pp. 269-282.
15. D. E. Littlewood, "Group characters and the structure of groups," *Proceedings of the London Mathematical Society* (2), **39** (1935), pp. 150-199.
16. M. Zia-ud-Din, "The characters of the symmetric group of order 11!," *Proceedings of the London Mathematical Society* (2), **39** (1935), pp. 200-204.
17. ———, "The characters of the symmetric group of degrees 12 and 13," *Proceedings of the London Mathematical Society* (2), **42** (1937), pp. 340-355.
18. C. Kostka, "Über den Zusammenhang zwischen einigen Formen von symmetrischen Funktionen," *Journal für die reine und ang. Mathematik* (Crelle), **93** (1882), pp. 89-123.
19. ———, "Tafeln und Formeln für symmetrische Funktionen," *Jahrbuch der Deutschen math. Ver.*, **16** (1907), pp. 429-451.
20. ———, "Tafeln für symmetrische Funktionen bis zur elften Dimension mit kurzen Erlaüterungen," Prog. (5), *kgl. Gym. u. Realgymn. Insterburg* (1908).
21. W. Specht, "Die irreduziblen Darstellungen der symmetrischen Gruppe," *Mathematische Zeitschrift*, **39** (1935), pp. 696-711.
22. E. Wigner, "On the consequences of the symmetry of the nuclear Hamiltonian on the spectroscopy of nuclei," *Phys. Rev.*, 2nd Ser. **51** (1937), pp. 106-119.
23. ———, "On the structure of nuclei beyond oxygen," *Phys. Rev.*, 2nd Ser. **51** (1937), pp. 947-958.

THE HEAVISIDE OPERATIONAL CALCULUS.*

By D. G. Bourgin and R. J. Duffin.

In its primary form the Heaviside calculus is concerned with the interpretation and application of functions of the operator p where p takes x^n into nx^{n-1}. Various representations are,[1] of course, possible. Heaviside's developments as well as the closely related work of Volterra [2] on permutable functions of the closed cycle, depend on series expansion of $F(p)$ and term by term interpretation according to the association $p^{-\nu} \doteq x^{\nu-1}/\Gamma(\nu)$. The particular representation used in this paper is that of the Laplace-Mellin integrals, namely, $F(p) \doteq f(x)$ [3] stands for

$$(1.1) \qquad F(p) = \int_{-\infty}^{\infty} e^{-xp} f(x)\, dx$$

$$(1.2)^5 \qquad \frac{f(x+0)+f(x-0)}{2}{}^4 = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{xp} F(p)\, dp.$$

This paper may be considered a study of some special results in the theory of these integrals.

The specific concerns of this work include the validation of the asymptotic expansion theorem of Heaviside for a wide class of functions and two theorems

* This joint work was completed in all essentials while the junior author, R. J. Duffin, was in the physics department of the University of Illinois. Received by the Editors January 9, 1936; Revised October 5, 1936, and March 6, 1937.

[1] H. Jeffreys, "Operational methods," *Cambridge Tracts*; T. C. Fry, *Annals of Mathematics*, vol. 34 (1921), p. 184; N. Wiener, *Mathematische Annalen*, vol. 95 (1926), p. 95; P. Levy, *Bulletin Mathematique de France*, vol. 1 (1926), p. 174.

[2] Volterra and Peres, *Leçons sur la Composition.*

[3] The notation is due to B. van der Pol, *Philosophical Magazine*, vol. 8 (1929), p. 801.

[4] In this article wherever $f(x)$ stands alone on one side of an equation, the meaning $(f(x+0)+f(x-0))/2$ is to be ascribed to it.

[5] Since there is no finite natural boundary for $F(p)$ in the present work, it is tacitly assumed that analytic continuation is used in the cut plane. For operational application such continuation is usually carried out by the principle of "permanence of form."

Operational interpretations may, of course, be developed for specific function classes by introducing simple closed circuits or Hankel or Pochhammer contours instead of the ordinate $\mathfrak{R}(p) = c$. Such interpretations, in the writers' opinion, are not strictly speaking of "Heaviside" type, in general, since the property of vanishing of the functions, thus defined, for negative real values is given up. Moreover the intimate relationship with the Fourier integral stressed in this paper, is lost.

which may be used to establish many of the formal identities in the literature of the Heaviside theory as well as certain extensions of that discipline for instance to a conjugate Heaviside calculus. The most interesting contributions are those connected with the development of certain reciprocal kernel relationships and the solutions of the Laplace integral equation.

Closely allied to the study of Eqs. 1. 1, 1. 2 is, in a sense, that of Fourier transforms; one writes $e^{-cx}f(x)$ in place of the usual $f(x)$ and $e^{cx}F(p)$ instead of $F(p)$, viz.

$$F(c+it)=\int_{-\infty}^{\infty}f(x)e^{-x(c+it)}dx \tag{1.11}$$

$$e^{-cx}f(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}F(c+it)e^{itx}dt. \tag{1.21}$$

These are precisely the integrals of Eqs. 1. 1 and 1. 2.

In general, Fourier integral theorems imply results for the Mellin Transformation. Evidently, then, conditions such as the bounded variation of $f(x)$ in the neighborhood of a point and the absolute integrability of $e^{-cx}f(x)$ over the axis of reals,[6] are sufficient for the inversion formulae, Eq. 1. 11 and Eq. 1. 21, provided that the Cauchy principal value[7] $L_{\lambda\to\infty}\int_{-\lambda}^{\lambda}\int_{-\infty}^{\infty}$ is understood in evaluating the infinite integrals.

In order to extend the set of operators, $f(x)$ in Eqs. 1. 1 and 1. 2 is assumed to vanish on the negative real axis and all "permissible" operators are such as to leave this function class invariant, (i. e. the lower limit of the integral of Eq. 1. 1 may always be taken as 0). This rules out, for instance,

---

[6] E. W. Hobson, *Functions of a Real Variable*, Cambridge Press, 2nd edition, vol. 2, p. 721. Throughout this paper the bounded variation condition introduced to guarantee the limit may be generally replaced by any other of the Fourier integral or Fourier Series conditions for convergence at a point.

[7] Some such condition is essential for the integral (with real $f(x)$)

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{ip(x-t)}f(x)\,dx dp$$

may be written

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x)\cos p(x-t)\,dxdp+i\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x)\sin p(x-t)\,dxdp.$$

The second integral on the right requires much stronger conditions than does the first for convergence. P. Pi Collega, *Mathematische Zeitschrift*, vol. 40 (1935), p. 349. However, the integrand is easily seen to be an odd function in p so that the Cauchy limit on p exists and is 0 for functions satisfying the condition for existence of the Fourier cosine integral.

the operator e^{hp} for $e^{hp}f(x)$ is $f(x+h)$ formally, which is non-zero in general for $(h>0)\ -h<x<0$. However, e^{-hp}, $h>0$ is a permissible operator.

We proceed now to establish in a direct fashion, an asymptotic expansion theorem of considerably greater content and precision than Heaviside's "rules."

THEOREM 1. (a) *$F(p)$ is analytic except for poles of order $\mu_1,\cdots,\mu_n$ at $p_1,\cdots,p_n$; $F(p)$ has essential singularities at $e_1,\cdots,e_t$. In each sufficiently small deleted neighborhood $F(p)$ is analytic and expansible in a Laurent series $\sum_0^\infty A_{kn}(p-e_k)^n+\sum_0^\infty B_{kn}(p-e_k)^{-n-1}$. $F(p)$ has branch points of finite order at $b_1,\cdots,b_m$ in the neighborhoods of which*

$$F(p)=(p-b_i)^{-\alpha_i}\psi_i(p-b_i)+\phi_i(p-b_i),\qquad \alpha_i>0;$$

the power series

$$\psi_i(p-b_i)=\Sigma\, a_{in}(p-b_i)^n,\qquad \phi_i(p-b_i)=\Sigma\, c_{in}(p-b_i)^n$$

converge for $|p-b_i|\leq r_i>0$. (b) *$L_{|p|\to\infty}|F(p)|\to 0$ uniformly for $\pi/2\leq\arg p-c\leq 3\pi/2$. If b is the abscissa of the singularity furthest right the ordering is such that $\mathcal{R}(p_j)=\mathcal{R}(b_i)=\mathcal{R}(e_k)=b$ for the first l values of j, the first q values of i, and the first f values of k.*

Under these hypotheses on $F(p)$

$$(2)\quad F(p)\doteq f(x)\sim\sum_{j=1}^{l}x^{\mu_j-1}e^{p_jx}\,\mathrm{Res.}\,F(p)(p-p_j)^{\mu_j-1}/\Gamma(\mu_j)$$
$$+\sum_{i=1}^{q}\sum_{0}^{\infty}e^{b_ix}a_{in}x^{\alpha_i-n-1}/\Gamma(\alpha_i-n)+\sum_{1}^{f}\sum_{0}^{\infty}e^{e_kx}B_{kn}x^n/\Gamma(n+1).$$

For $x>0$ we use the closed contour made up of the ordinate $\mathcal{R}(p)=c>b$, the left hand infinite semi-circle on this ordinate together with the necessary non-intersecting branch cuts and small circles about the poles. By Cauchy's Theorem

$$(2.1)\quad \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}e^{xp}F(p)\,dp-\Sigma\ (\text{resides}+\text{integrals about essential singularities}+\text{integrals around branch cuts}+\text{integrals on the semi-circle})=0.$$

The residues evidently contribute just the terms involved in the first summation in the theorem.

Surround the essential singularities by non-intersecting circles of radii s_k lying within the regions of convergence of the Laurent series. Because of the

uniform convergence of these series on the circles term by term integration is justified. The inverse powers alone contribute to $f(x)$ and their effect is epitomized in the third group of terms of Eq. 2. Evidently

$$L_{n\to\infty} | B_{kn} |^{1/n} = o(s_k).$$

The function $\sum_0^\infty B_{kn}x^n/\Gamma(n+1)$ is then an entire function of minimal type, and thus is dominated, for large x by $Ae^{\epsilon x}$, ϵ arbitrarily small > 0. The minimal property indicates that the contributions of e_{f+w}, $(w = 1, \cdots, t - f)$, are to be compared according to the value of the exponentials $e^{-x(b-e_k)}$. For $k = f + w$ these are negligible for large x, and accordingly these terms as well as those arising from $p_{l+1}, \cdots, p_n$ are unimportant.

Consider now the third term under the summation sign in Eq. 2. 1. The branch cut for b_i may be taken as a straight line inclined at an angle ϑ with the real axis $3\pi/2 > \vartheta > \pi/2$. For ease of exposition alone, ϑ will be taken as π in the work below.[8] The contour around the cut may be taken as made up of the part of the upper and lower edges of the cut terminating to the left of b_i within the circle of radius r_i and a loop, denoted hereafter by C_i, around the branch point and lying entirely within this circle to complete the cycle. The transformation $z = p - b_i$ brings the branch point b_i to the origin and introduces the factor $e^{b_i x}$. On writing $-\infty \leqq \mathcal{R}(z) \leqq -\rho_i > -r_i$ with $\arg z = \pi i$ or $-\pi i$ for the upper and lower boundaries of the cut, the absolute value of the contribution due to these parts of the dissected contour may be exhibited as

$$(2.2) \quad \left| \frac{1}{2\pi i} e^{b_i x} \int_{\rho_i}^{\infty} e^{-rx}[F(re^{i\pi} + b_i) - F(re^{-i\pi} + b_i)]dr \right| \leqq K_1 e^{(b_i-\rho_i)x}/x$$

since $F(p)$ is bounded on the cut away from the branch point by $2\pi K_1$. This is later shown to be negligible in comparison with the other terms in the final developments so that the *behavior of $F(p)$ in the immediate neighborhood* of the branch points determines the result.

For the term $z^{n-a_i}a_{in}$ the loop integral around the branch point may be written, on making the substitution $zx = -u$, as

$$(2.3) \qquad \frac{a_{in}}{2\pi i} e^{b_i x} x^{a_i-n-1} \int_{C'_i} e^{-u}(-u)^{n-a_i} du$$

where C'_i is the Hankel loop starting from the point on the upper edge of the cut (in the u plane) of abscissa $\rho_i x$ and passing counter clockwise about the

[8] Indentations to avoid possible singularities are tacitly neglected since their sole effect is at most a change in K_1 of Equation 2. 2.

origin to the point just below on the lower edge. For x large enough this approaches [9]

(2.31) $$\frac{a_{in}}{\pi} x^{\alpha_i-n-1} \sin \alpha_i\pi \, e^{b_ix}\, \Gamma(n+1-\alpha_i)(-1)^n.$$

For identification as a term in the second expansion in Eq. 2 one need remark here and later that

$$\frac{\sin \alpha_i\pi}{\pi}(-1)^n = \frac{\sin(\alpha_i - n\pi)}{\pi} = [\Gamma(1-(\alpha_i-n))\Gamma(\alpha_i-n)]^{-1}.$$

The maximum difference between Eq. 2.31 and Eq. 2.3 is found, essentially, by bounding $\int_\rho^\infty e^{-rx} r^{n-\alpha_i} dr$. For sufficiently large x this difference is then easily shown to be inferior to

(2.32) $$K_2 x^\mu e^{(b_i-\rho_i)x}.$$

We wish to show now that the series defined by the sum of terms of Eq. 2.31 exclusive of the e^{b_ix} factor, is an asymptotic series, namely that

(2.4) $$L_{x\to\infty} x^{-\alpha_i+M+1} \left| \frac{1}{2\pi i}\int_{C_i} e^{(zx)}\psi_i(z) z^{-\alpha_i} dz - \frac{1}{\pi}\sum_0^M a_{im} \sin \alpha_i\pi\Gamma(n+1-\alpha_i) x^{\alpha_i-n-1}(-1)^n \right| \to 0$$

where $\psi_i(z)z^{-\alpha_i}$ is written instead of $F(z+b_i)$ since evidently the loop integral for $\phi_i(z)$ is 0.

The expression under the absolute value signs in Eq. 2.4 is surely inferior to

(2.5) $$\left| \frac{1}{2\pi i}\int_{C_i} e^{zx} z^{-\alpha_i+M+1}\bar{\psi}_i(z)\, dz \right| + K_2 x^\mu e^{-\rho_i x} + K_1 e^{(b_i-\rho_i)x}/x$$

where $\bar{\psi}_i(z) z^{M+1}$ is the remainder after subtracting off the first $M+1$ terms of the series expansion of $\psi_i(z)$. Our immediate problem is to show that Eq. 2.5 is at most $o(x^{\alpha_i-M-1})$ for $x \to \infty$. Hence the last terms in that expression is unimportant.

Evidently $\bar{\psi}(z)$ is analytic for $|z| \leq \rho_i$ and so $\bar{\psi}(z) < K_3$. We may deform the loop into a circle $|z| = \sigma_i/x$, $\sigma_i < \rho_i$ and the upper and lower edges of the cut in the range $-\rho_i \leq \mathcal{R}(z) \leq -\sigma_i/x$.

The contribution from the integrals along the cut is less than

[9] Whittaker and Watson, *Modern Analysis*, 3rd edition, p. 244.

$$(2.6)\qquad \left| \frac{\sin \alpha_i \pi}{\pi} x^{a_i - M - 2} K_3 \int_{\sigma_i}^{\rho_i x} e^{-r} r^{-a_i + M + 1} dr \right| \leq K_3 \left| x^{a_i - M - 2} \int_{\sigma_i}^{\infty} e^{-r} r^{-a_i + M + 1} dr \right| \leq K_4 \mid x^{a_i - M - 2} \mid .$$

Similarly the integral around the small circle is inferior to

$$(2.61)\qquad K_3 \sigma^{-a_i + M + 2} x^{a_i - M - 2}.$$

The dominants found in Eqs. 2. 6 and 2. 61 guarantee that Eq. 2. 5 is at most $o(x^{a_i - M - 1})$ for $x \to \infty$. The asymptotic character of the expansions considered has thus been established.

According to a simple extension of a classical lemma due to Jordan the integral on the infinite semi-circle vanishes when account is taken of (b). Furthermore, since the dominants given in Eqs. 2. 2 and 2. 22 involve an exponential decrease faster than that of terms with an $e^{b_i x}$ factor, it follows that the value of the Mellin integral in Eq. 2. 1 is asymptotically approximated by the expansions in the statement of the theorem. The assurance that the Mellin integral really exists follows on the observation that the terms under the summation sign in Eq. 2. 1 remain finite for a divergent sequence of sufficiently large semi circles erected on $\mathfrak{R}(p) = c$ and that Eq. 2. 1 is valid for each of the resulting closed contours.

The proof is complete.[10]

If $F(p)$ is restricted to correspond to a real function $f(x)$ (of the real variable x) then a_{in} and B_{kn} are real and each exponential of imaginary argument is replaced by a sine or cosine. This follows immediately on remarking that then

$$\mathfrak{R}F(p) = \mathfrak{R}F(\bar{p}); \qquad \mathfrak{I}F(p) = -\mathfrak{I}F(\bar{p})$$

which imply that b_i, e_i and p_i occur in conjugate pairs.

Volterra composition, namely

$$F_1(p)F_2(p) \doteq \int_0^x f_1(z) f_2(x-z)\, dz$$

[10] This clears up the doubts arising in the minds of some of the Heaviside followers, for instance those of Carson, concerning the validity and applicability of the asymptotic Heaviside expansions. Cf. Carson, *Electrical Circuits*, p. 84. Carson's difficulties are but imperfectly answered in Levy's paper, *ibid.*, and the main point is not touched. The situation is, of course, summarized in Equation 2. March, *Bulletin of the Mathematical Society*, vol. 33 (1927), p. 311, has utilized a similar method for a rather restricted case. He does not, moreover, give any exact sufficient conditions for validity and omits the demonstration of the asymptotic property of the development.

follows under suitable restrictions from Eqs. 1. 1 and 1. 2 when it is recalled that $f_1(x)$ and $f_2(x)$ vanish for negative values of their arguments. Here we are interested in the (inverse) composition on the p functions, namely

$$(3) \qquad F_3(p) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} F_1(p-z)F_2(z)\,dz = \int_0^\infty f_1(x)f_2(x)e^{-px}dx \doteq f_1(x)f_2(x).$$

THEOREM 2. *If $f_1(x)e^{-c_1x}$, $f_2(x)e^{-c_2x}$ belong to $L_2(0,\infty)$,*

$$\text{(b)} \qquad \Re(p) = d > c + c_2, \qquad c > c_1 \geqq c_2$$

then[11] *Eq. 3 is valid and $F_3(p)$ is analytic for $\Re(p)$ satisfying* (b) *is at worst $o(1)$ for $|q| \to \infty$, $q = \Im(p)$ and the Mellin integral with $F_3(p)$ is summable $C\,1$ to $f_1(x)f_2(x)$*[4] *wherever this has meaning.*

The proof is immediate, for it

$$(3.1) \qquad F_i(\sigma + it) = \text{l.i.m.} \int_0^\infty f_i(x)e^{-x(\sigma+it)}dx \ \text{ with } \sigma > c_i$$

it is well known that $F_i(\sigma+it)$ is not only analytic in the half plane $\sigma > c_i$ but that

$$(3.2)^{12} \qquad \int_{\sigma-i\infty}^{\sigma+i\infty} |F(z)|^2\,dt < \infty.$$

This guarantees that $F_1(c+it)$ and $F_2(d+iq-(c+it))$ are Fourier transforms of class L_2 on the ordinates provided $c > c_1$ $d - c > c_2$. It is easy to see that the analogue of the Parseval identity becomes[13]

$$(3.3) \qquad L_{A\to\infty}\frac{1}{2\pi i}\int_{-A}^{A} F_1(c+it)F_2(d+iq-(c+it))\,dt = L_{A\to\infty}\frac{1}{2\pi i}\int_{c-iA}^{c+iA} F_1(z)F_2(p-z)\,dz = L_{B\to\infty}\int_0^B f_1(x)f_2(x)e^{-xp}dx.$$

The first integrand evidently belongs to L_1 by the Schwarz Inequality. The

---

[11] Somewhat similar theorems, involving different conditions, in connection with Dirichlet Series, occur in the recent literature. Cf. D. V. Widder, *American Journal of Mathematics*, vol. 49 (1927), p. 321, for the case $f_i(x)$ absolutely integrable; V. Bernstein, *Series Dirichlet*, Appendix I for the case $f_i(x)$ analytic in sectors.

[12] Paley and Wiener, "Fourier transforms," *American Colloquium Publications* (hereafter designated P. W.) Theorem 5.

[13] For instance by paralleling the steps in N. Wiener, *Acta Mathematica*, vol. 118 (1930), p. 55.

last integral may be written $\int_0^\infty g(x)e^{-x(h+iq)}dx$, $g(x) \subset L_1(0, \infty)$ where $h = d - c_1 - c_2 > 0$ and $g(x)$ is the product of two functions each of which belongs to $L_2(o, \infty)$

$$g(x) = (e^{-xc_1}f_1(x))(e^{-c_2x}f_2(x)). \tag{3.4}$$

Accordingly, the integrals in Eq. 3. 3 not only exist but, by Eq. 3. 4, define an analytic function[14] $F_3(p)$ in the half plane determined by (b). In the event that the first integrand of Eq. 3. 3 or $g(x)$ also belongs to L_2 then, of course, $F_3(p)$ belongs to L_2 as well.

The deduction $L_{|q|\to\infty} |F_3(p)| = o(1)$, $\mathcal{R}(p) \geq d$ is correct whenever $f_3(x)$ (here $g(x)$) of Eq. 1. 1 belongs to $L_1(0, \infty)$, viz:

$$\left| \int_0^\infty g(x)e^{-x(h+iq)}dx \right| \leq \left| \int_0^A \right| + \left| \int_A^{A^{-1}} g(x)e^{-x(h+iq)}dx \right| + \left| \int_{A^{-1}}^\infty \right| .$$

For A sufficiently small the moduli of the first and last integrals on the right are inferior to ϵ uniformly in q. The second integral goes to 0, by the Riemann-Lebesgue lemma, for $|q| \to \infty$.

The remarks in the introduction make it clear that C 1 summability follows by direct extension from the known Fourier integral result since $g(x) \subset L_1$.

On making use of the relation $e^{-\lambda p}F(p) = f(x-\lambda)$ we may write Eq. 3 in a form convenient for many applications.

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{-xz}F_1(p-z)F_2(z)\,dz \tag{3.01}$$
$$= \int_0^\infty e^{-p\lambda}f_1(\lambda)f_2(\lambda-x)\,d\lambda = \int_x^\infty e^{-p\lambda}f_1(\lambda)f_2(\lambda-x)\,d\lambda.$$

The analyticity of $F_i(p)$ in the right half plane indicates that the singularities of $F_1(p-z)$ lie to the right of $\mathcal{R}(z) = c$ and those of $F_2(z)$ to the left. In the special case of polar singularities and $F_2(z)F_1(p-z) = o(z^{-1})$ uniformly in $\arg z - c$ on either the right or left infinite semi-circle constructed on the diameter $\mathcal{R}(z) = c$ for instance it is possible to close the contour at infinity and to contract in such wise as not to pass any singularities included in the interior. Thus the closed contour contains all the polar singularities of just one of the functions involved.[15]

---

[14] S. Bochner, *Vorlesungen über Four. Int.*, Leipzig, p. 145.

[15] This special case comprehends the usual Heaviside rules. $f_1(x) = 1$, e^{-ax}, x^n (n integral) leads to the Cauchy formula, to $F_2(p+a)$ and $(-d/dp)^n$ respectively. For non-integral n in the last example Theorem 2 provides the basis for a theory of fractional differentiation and integration of p functions comparable to that known for the x functions. In this connection compare Equation 5. 1.

The following simple identity may be made the basis for many novel developments as well as for a number of formal results already in the Heaviside literature.

$$(4)^{16} \qquad \int_0^\infty e^{-\rho_1 x} f_1(x) F_2(\rho_2 + x)\, dx = \int_0^\infty e^{-\rho_2 x} f_2(x) F_1(\rho_1 + x)\, dx.$$

It is convenient to use $\phi_i(z) = e^{-\rho_i z} f_i(z)$. Eq. 4 is then obviously valid when

$$(4.01) \qquad \int_0^\infty \int_0^\infty \phi_1(x)\phi_2(\lambda) e^{-x\lambda} dx d\lambda = \int_0^\infty \int_0^\infty \phi_1(x)\phi_2(\lambda) e^{-x\lambda} d\lambda dx.$$

The indicated inversion is justified for hypotheses such as [17]

THEOREM 3. *$\phi_i(z)$ is integrable L_1 over any finite closed range of positive z values not including 0 or ∞ and either side of Eq. 4 exists for absolute values of the integrand.*

A simple extension of Titchmarsh's theorem [18] 2. 62 suffices.

THEOREM 3A. $\int_0^\infty |\phi_1(x)|\, dx$, $\int_0^\Lambda |\phi_2(\lambda)|\, d\lambda$ $0 < \Lambda < \infty$, $\int_0^\infty \phi_i(\lambda)\, d\lambda$ $(i = 1, 2)$, *exist as Riemann integrals.*[19]

---

[16] $f_i(x)$ and ρ_i are treated as real for the proofs below. However, the results are valid generally, on splitting up the integrand into the four combinations

$$\mathfrak{R}(\mathfrak{I})\phi_1(x) \times \mathfrak{R}(\mathfrak{I})\phi_2(\lambda),$$

if the hypotheses are satisfied for the separate products. The normal case in operational theory involves only ρ_i complex in which case Theorem 3A alone is affected.

[17] The van der Pol result $\int_0^\infty F(p)\, dp = \int_0^\infty f(x)/x\, dx$ is the special case for which one of the functions is 1 and $\rho_1 = \rho_2 = 0$. This identity has been of extraordinary utility in the work of B. van der Pol, *loc. cit.* and later *Philosophical Magazine* papers. The Riemann-Stieltjes equation

$$h(y) = \int_0^\infty F(p) e^{-py} dp = \int_0^\infty f(x)/x + y\, dx$$

arises on taking one of the ϕ's as e^{-hy}.

[18] E. C. Titchmarsh, *Theory of Functions*, Oxford Press. (Hereafter designated T.).

[19] This theorem admits cases such as $\phi_2(x) = xe^{e^x}\sin e^{e^x}$. For this function $\int_0^\infty e^{-xp}\phi_2(x)\, dx$ converges conditionally, only, for $p \geqq 0$. Eq. 1. 2 with $F_2(p)$ requires generally a summability interpretation. For summability C 1 the validation follows from a result of Hardy's, G. H. Hardy, *Messenger of Mathematics*, vol. 47 (1917), p. 178. Since the central inequality, Eq. 4. 2, does not require uniform convergence, it is evident that the hypotheses of Theorem 3 and 3A may be further considerably weakened.

The proof is straightforward. We assert that

$$(4.1)\qquad \int_0^X \int_0^\Lambda \phi_1(x)\phi_2(\lambda)e^{-x\lambda}dx d\lambda = \int_0^\Lambda \int_0^X \phi_1(x)\phi_2(\lambda)e^{-x\lambda}d\lambda dx.$$

Under our hypotheses both iterated integrals exist for absolute values of the integrand. Hence the integrations may be interpreted in the sense of Lebesgue, but then the integration order is immaterial [20] and this result must hold also for the Riemann interpretation.[21]

After a preliminary integration by parts, it is easy to demonstrate the uniform convergence of $\int_0^\infty \phi_2(\lambda)e^{-x\lambda}d\lambda$ for $x \geqq 0$. Accordingly

$$\left| \int_0^X \phi_1(x) \int_\Lambda^\infty \phi_2(\lambda)e^{-x\lambda}d\lambda dx \right| \leqq \int_0^X |\phi_1(x)|\, dx \left| \int_\Lambda^\infty \phi_2(\lambda)e^{-x\lambda}d\lambda \right|$$
$$\leqq \epsilon_1 \int_0^\infty |\phi_1(x)|\, dx < \eta_1$$

for (a) Λ fixed $\geqq \Lambda_1$ and all X, (b) X fixed and all $\Lambda \geqq \Lambda_2$.

The absolute integrability of $\phi_2(\lambda)$ over finite ranges justifies the assertion that for fixed Λ, X_1 exists such that

$$(4.3)\quad \left| \int_0^\Lambda \phi_2(\lambda)d\lambda \int_X^\infty \phi_1(x)e^{-x\lambda}dx \right| \leqq \epsilon_2 \int_0^\Lambda |\phi_2(\lambda)|\, d\lambda \leqq \eta_2 \text{ for all } X \geqq X_1.$$

Eqs. 4. 1, 4. 2 and 4. 3 are sufficient to establish the validity of the change in integration order [22] involved in Eq. 4. An obvious generalization to cover the case that $\int_l^\Lambda |\phi_2(\lambda)|\, d\lambda$ exists for $l > 0$, $\Lambda < \infty$ only is included by interchanging the rôles of 0 and ∞ in the above proof.

THEOREM 3B. $\phi_i(z)$ belongs to $L_k(0, \infty)$, $1 < k \leqq 2$.

The first half of the conditions of Theorem 3 are easily shown to be satisfied for integrability L_k implies integrability L_1 over finite ranges. We show now that the last condition of Theorem 3 is also met.

If $\psi_i(z) = \int_0^\infty e^{-xz} |\phi_i(x)|\, dx$ it is known [23] that

$$\int_0^\infty |\psi_i(z)|^{k'}dz < c\left(\int_0^\infty |\phi_i(z)|^k dz\right)^{k'/k}, \qquad c < \infty$$

[20] T., Theorem 12. 6.

[21] T., p. 340.

[22] W. H. Young, *Cambridge Philosophical Transactions*, vol. 21 (1910), p. 48.

[23] G. H. Hardy, *Journal of the London Mathematical Society*, vol. 8 (1933), p. 114.

and $1/k + 1/k' = 1$. It may easily be shown that $\psi_i(z)$ is continuous for $z > 0$, hence $\psi_i(z)$ is measurable and therefore $\psi_i(z)$ belongs to $L_{k'}(0, \infty)$.

By Holder's inequality

$$\int_0^\infty |\psi_1(z)\phi_2(z)|\,dz \leq \left(\int_0^\infty |\psi_2(z)|^{k'}dz\right)^{1/k'}\left(\int_0^\infty |\phi_1(z)|^{k}\right)^{1/k}$$
$$\leq C\left(\int_0^\infty |\phi_2(z)|^{k}dz\right)^{1/k}\left(\int_0^\infty |\phi_1(z)|^{k}dz\right)^{1/k} < \infty.$$

Thus the sufficient hypotheses of Theorem 3 are implied by those of Theorem 3B.

As a first direct application of Theorems 2 and 3, we present an independently interesting development of a calculus, conjugate, in a certain sense, to that of Heaviside. Negative integral powers of the variable are not comprehended [24] by our representation (Eq. 1.1 and Eq. 1.2) of the Heaviside calculus. However, an operational theory can be stated for such functions on interchanging the meaning of operator and variable in Eqs. 1.1 and 1.2 so that q and x correspond to the previous x, p respectively. The new variable (x) now ranges over the entire complex plane. An operational expression may be interpreted simultaneously according to both the Heaviside and the conjugate symbolic theories on decomposing the operand

$$\psi(p)g(x) = \psi(p)f(x) + \psi(q)F(x)$$

where $g(x) = f(x) + F(x)$ and the notation for the functions indicates the sets to which $f(x)$ and $F(x)$ belong. This decomposition is no longer unique when fractional powers are present for a distinction regarding domain of consideration must be made for $x^{-\nu}$, $0 < \nu < 1$, for instance, accordingly as it is included in the Heaviside set or the set of the conjugate theory functions.

The following algorithms are straightforward consequences of Theorem 3 for functions fulfilling the restrictions stated there.

$$qF(x) = \frac{d}{d(-x)}F(x) \qquad \frac{d}{dq}f(q) \doteq (x)F(x)$$

$$\text{(a)} \qquad q^{-1}F(x) = \left(\int_0^\infty e^{-x\lambda}f(\lambda)/\lambda d\lambda\right) = \int_0^\infty F(x+\lambda)\,d\lambda = \int_x^\infty F(t)\,dt.$$

(From the viewpoint of Volterra composition there is a formal analogy here to the usual $p^{-1}f(x)$ interpretation

$$\text{i. e. } p^{-1}f(x) = \int_0^\infty f(t)h(x-t)\,dt\left(= \int_0^X f(x)\,dx\right)$$

[24] Formally $p \log p = -x^{-1}$, but this is not in the domain of Equation 1.1.

with $h(y)$ the unit function vanishing on the negative axis of reals. For the special case of the conjugate theory cf. (a) when x is real, the same composition formula may be considered to apply, but $h(y)$ now represents the unit function for negative real values. The composition property with the reflected unit function holds throughout the conjugate theory when x is real. Vide (b) and (c) below.

$$(5.1)\quad \begin{aligned} &\text{(b)}\quad (q+c)^{-1}F(x) = \int_0^\infty e^{-c\lambda}F(x+\lambda)\,d\lambda = \int_x^\infty e^{-c(t-x)}F(t)\,dt \\ &\text{(c)}\quad (q+c)^{-n}F(x) = \left(\int_0^\infty e^{-x\lambda}f(\lambda)/(\lambda+c)^n d\lambda\right) \\ &\qquad\qquad = \int_0^\infty e^{-c\lambda}\lambda^{n-1}F(\lambda+x)/\Gamma(n)\,d\lambda \end{aligned}$$

The last result may be obtained operationally (for n a positive integer) by formally differentiating $(q+c)^{-1}F(x)$ with respect to $-c$. With $c=0$ Eq. 5.1 may be interpreted as a fractional integral of a p function.

Theorem III determines the function associated with $f(x^n)$ at least for n integral. For $n=2$ we start with

$$f_1(x) = \exp(-\lambda^2/x)/4(\pi x)^{1/2} \doteq F_1(p) = \exp(-\lambda p^{1/2})/p^{1/2};$$

whence quite directly

$$4\pi^{1/2}f_2(x^2) \doteq \int_0^\infty F_2(z^2)\exp(-p^2)/z^2 dz.$$

The application to constant coefficient differential equations is direct and may be briefly summarized. Consider the differential equation with constant coefficients

$$(5.2)\quad \begin{aligned} &\sum_{i=0}^{n} \alpha_i y^{n-i}(x) = L(d/dx)y = L(-q)y = F(x); \\ &\lim_{x\to\infty} y,\cdots,y^{n-1}\to 0, \qquad \mathcal{R}(x)\leq c. \end{aligned}$$

We may write, on the assumption of no positive real zeros,

$$(5.3)\qquad y(x) = \int_0^\infty e^{-x\lambda}f(\lambda)/L(-\lambda)\,d\lambda,\ \text{where}\ f(\lambda)\doteq F(x).$$

In accordance with Theorem 3 (or Eq. 5.1) this may be expressed as

$$(5.4)\qquad y(x) = \int_0^\infty \frac{dA(\lambda)}{d\lambda}F(x+\lambda)\,d\lambda$$

where $A(\lambda)$ is the ordinary "indicial" function on the Heaviside theory corresponding to $L(-\lambda)$. It is well known in fact that

$$A(\lambda) = \frac{1}{L(0)} + \sum e^{a_j\lambda}/\alpha_j \frac{d}{dq} L(-q)|_{a_j}, \qquad L(-\alpha_j) = 0$$

if the zeros are distinct and not positive real.

However, another viewpoint may be used in connection with Eq. 5. 3. Define

$$\psi(\alpha_j x) \equiv e^{-a_j x} \int_{-a_j x}^{\infty - \mathcal{J}(a_j x)} e^{-v}/v \, dv, \qquad \mathcal{J}(v) \equiv \mathcal{J}(-\alpha_j x)$$

then for the case $F(x) = x^{-1}$ Eq. 5. 3 reduces to

$$y(x) \equiv B(x) = \sum_j \psi(\alpha_j x) / \frac{d}{dq} L(-q)|_{a_j}$$

for distinct not positive real zeros of $L(-p)$.

This is the analogue of the Heaviside-Carson "indicial" function's derivative. For the equation with a general $F(x)$ the inverse composition process of Theorem 2 gives, as an alternative to the formula of Eq. 5. 4, the solution

$$\text{(5. 41)} \qquad y(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} B(t) F(x-t) \, dt.$$

The more usual formulation of the Mellin integrals is connected with the operator $s = xd/dx$ and is expressed

$$\text{(1. 1 bis)} \qquad G(s) = \int_0^\infty v^{s-1} g(v) \, dv$$

$$\text{(1. 2 bis)} \qquad 2\pi i \, g(v) = \int_{d-i\infty}^{d+i\infty} v^{-s} G(s) \, ds.$$

Combination of the operators s and p may well be expected to have special interest in operational theory. Consider then

$$\text{(6)} \qquad \gamma(s, p) = \int_0^\infty v^{s-1} e^{-vp} \phi(v) \, dv.$$

Some striking inversion relations arise through the intermediation of Eq. 6. We write

$$\text{(6. 1)} \qquad \psi(s) = \int_{A_1}^{A_2} \gamma(s, \lambda) f(\lambda) \, d\lambda.$$

The nature of the reciprocal formula is indicated by the following purely formal developments

$$\text{(6. 11)} \qquad \psi(s) = \int_{A_1}^{A_2} \int_0^\infty v^{s-1} e^{-v\lambda} \phi(v) f(\lambda) \, dv d\lambda = \int_0^\infty v^{s-1} \phi(v) F(v) \, dv$$

if Theorem 3 applies, where

$$F(v)=\int_{A_1}^{A_2} e^{-xv}f(x)\,dx.$$

Then

$$\phi(v)F(v)=(1/2\pi i)\int_{d-i\infty}^{d+i\infty} v^{-s}\psi(s)\,ds$$

and

$$f(x)=(1/2\pi i)^2\int_{c-i\infty}^{c+i\infty}\int_{d-i\infty}^{d+i\infty} e^{xv}v^{-s}\psi(s)/\phi(v)\,ds\,dv$$

$$(6.2)\qquad =(1/2\pi i)\int_{d-i\infty}^{d+i\infty} K(s,x)\psi(s)\,ds$$

with

$$K(s,x)=(1/2\pi i)\int_{c-i\infty}^{c+i\infty} e^{xv}v^{-s}/\phi(v)\,dv.^{25}$$

The rigorous validation of this mode of derivation of the important Eq. 6.2 presents difficulties because the integrals involved are generally not absolutely convergent.[26] For the case $\phi(v)=(1-e^{-v})^{-1}$,

$$\gamma(s,v)=\Gamma(s)\zeta(s,v)\quad\text{where}\quad \zeta(s,v)=\sum_{n=0}^{\infty}(v+n)^{-s}$$

is the generalized zeta-function, Eq. 6.2 takes the elegant form

$$(6.21)\quad f(x)=(1/2\pi i)\int_{d-i\infty}^{d+i\infty}[x^{s-1}-\mathbf{1}(x-1)^{s-1}]\psi(s)\,ds$$
$$\mathbf{1}(z)=0 \text{ for } z<0,=1 \text{ for } z>0, \text{ and } =1/2 \text{ for } z=0.$$

THEOREM 4. *Eq. 6.21 is valid if*

$$x,x-1,\cdots,x-r\ (x-r\geqq A_1>x-r-1)$$

are interior to neighborhoods of bounded variation of $f(x)$ *and* $A_1>0$, $A_2=\infty$ *and furthermore that* $f(x)$ *belongs to* $L_1(A_1,\infty)$, *provided that* $\mathcal{R}(s)=d>1$.

Consider

$$(6.3)\quad (1/2\pi i)\int_{d-i\beta}^{d+i\beta} x^{s-1}\psi(s)\,ds=(1/2\pi i)\int_{d-i\beta}^{d+i\beta} x^{s-1}\int_{A_1}^{\infty} f(\lambda)\zeta(s,\lambda)\,d\lambda\,ds$$
$$=(1/2\pi i)\int_{A_1}^{\infty} f(\lambda)\int_{d-i\beta}^{d+i\beta} x^{s-1}\zeta(s,\lambda)\,ds\,d\lambda$$
$$=(1/2\pi i)\int_{A_1}^{\infty} f(\lambda)\int_{d-i\beta}^{d+i\beta}\sum_{0}^{\infty}(\lambda+n)^{-s}x^{s-1}\,ds\,d\lambda.$$

The inversion of order of integration is justified by the observation that for

[25] If $\phi(v)=1/\sum_{0}^{\infty}A_k e^{-B_k v}$ then $K(s,x)=\sum_{0}^{\infty}A_k\cdot(x-B_k)\mathbf{1}(x-B_k)$ formally.

[26] In fact $L_{|t|\to\infty}\,\gamma(d+it,p)$ may not exist. K. Ananda-Rau, *Proceedings of the London Mathematical Society*, vol. 19 (1920), p. 114.

$\mathcal{R}(s) = d > 1$, $|\zeta(s,\lambda)|$ is continuous in λ and s when $A_1 \leq \lambda \leq \infty$ and accordingly the integrals are absolutely convergent.

On carrying out the integration there results

$$\int_{A_1}^{\infty} f(\lambda) \sum_0^{\infty} \frac{(x/n+\lambda)^d}{\pi x \log (x/n+\lambda)} \sin \beta \log (x/n+\lambda)\, d\lambda. \tag{6.31}$$

The term by term integration is correct for $\sum_0^{\infty} x^{-1}(x/n+\lambda)^s$ is uniformly convergent (in $\mathcal{I}(s)$) for $\mathcal{R}(s) = d > 1$, $\beta < \infty$.

We may write

$$\left| \sum_{N_x}^{\infty} \frac{(x/n+\lambda)^d}{\pi x \log (x/n+\lambda)} \sin (\beta \log (x/n+\lambda)) \right| < \epsilon_1$$

for N_x $(> x)$ sufficiently large, uniformly in β and λ. Thus

$$\int_{A_1}^{\infty} |f(\lambda)| \left| \sum_{N_x}^{\infty} \frac{(x/n+\lambda)^d}{\pi x \log (x/n+\lambda)} \sin (\beta \log x/n+\lambda) \right| d\lambda < \epsilon_2. \tag{6.4}$$

Since also

$$L_{A\to\infty} \sum_0^{N'} \int_A^{\infty} \to 0 \tag{6.41}$$

uniformly in β for finite N', x we may invert the operations in Eq. 6.31 to get

$$\sum_0^{\infty} \int_{A_1}^{\infty} f(\lambda) \frac{(x/n+\lambda)^d}{\pi x \log (x/n+\lambda)} \sin (\beta \log x/n+\lambda)\, d\lambda$$

or what is essentially the same thing according to Eq. 6.4

$$\sum_0^{N_x} \int_{A_1}^{\infty} f(\lambda) \frac{(x/n+\lambda)^d}{\pi x \log (x/n+\lambda)} \sin (\beta \log x/n+\lambda)\, d\lambda. \tag{6.5}$$

Accordingly,

$$\begin{aligned} &L_{\beta\to\infty} (1/2\pi i) \int_{d-i\beta}^{d+i\beta} x^{s-1}\psi(s)\, ds \\ &\quad = L_{\beta\to\infty} \sum_0^{N_x} \int_{A_1}^{\infty} f(\lambda) \frac{(x/n+\lambda)^d}{\pi x \log (x/n+\lambda)} \sin (\beta \log x/n+\lambda)\, d\lambda \\ &\quad = L_{\beta\to\infty} \sum_0^{N_x} \int_{\log x/n+A_1}^{-\infty} e^{(d-1)z} f(xe^{-z}-n) \frac{\sin \beta z}{\pi z}\, dz \end{aligned} \tag{6.32}$$

$$\begin{aligned} &\int_{\log x/n+A_1}^{-\infty} | e^{(d-1)z} f(xe^{-z}-n) |\, dz \\ &\quad = \int_{A_1}^{\infty} | x^{d-1} f(\lambda)/(n+\lambda)^d |\, d\lambda < \infty. \end{aligned} \tag{6.6}$$

Because of Eq. 6.6 and Eq. 6.41 the Riemann-Lebesgue lemma may be applied to show that the Dirichlet integrals in Eq. 6.32 vanish unless $x \geq n + A_1$. Accordingly, the limit of the first integral in Eq. 6.32 is

$$f(x) + f(x-1) + \cdot \cdot \cdot f(x-r) \tag{6.7}$$

where $x - r \geqq A_1 > x - r - 1$, provided the function is of bounded variation in the neighborhoods of the arguments in question.

Similarly

$$L_{\beta\to\infty} (1/2\pi i) \int_{d-i\beta}^{d+i\beta} 1(x-1)(x-1)^{s-1}\psi(s)\,ds = f(x-1) + \cdots f(x-r) \tag{6.8}$$

Subtraction of the expressions in Eq. 6.8 from those in Eq. 6.7 yields the desired result.

THEOREM 4A. *If $f(x) \subset L_1(A_1, \infty)$, Eq. 6.21 is valid in the sense that the right-hand integral is summable* (C, 1) *wherever $f(x+0) + f(x-0)$ exists.*

We have merely to investigate

$$L_{\beta\to\infty} (1/2\pi i) \int_{d-i\beta}^{d+i\beta} \left(1 - \frac{|\mathfrak{J}(s)|}{\beta}\right) (x^{s-1} - 1(x-1)(x-1)^{s-1})\psi(s)\,ds.$$

The steps and reasoning of the argument are precisely the same in detail as in the proof above. The only change is that the resulting integrals (Eq. 6.32 for instance) are of Fejer instead of Dirichlet type.

One immediate application is furnished by the Laplace integral equation, Eq. 1.1 where p is now a real variable, $F(p)$ is supposed known for $p \geqq p_0$ and $f(x)$ is required. There is no fundamental restriction in assuming $f(x)$ bounded and of class $L_1(0, \infty)$.[27]

For $f(x)$ satisfying the conditions of Theorems 4 or 4A we may exhibit solutions in the form [28]

$$f(x) = (1/2\pi i) \int_{d-i\infty}^{d+i\infty} (x^{s-1} - 1(x-1)(x-1)^{s-1}) \int_0^\infty (p^{s-1}F(p)/\Gamma(s)(1-e^{-p}))\,dp\,ds \tag{7}$$

or

$$= L_{\beta\to\infty} (1/2\pi i) \int_{d+i\beta}^{d-i\beta} \left(1 - \frac{|\mathfrak{J}(s)|}{\beta}\right)(x^{s-1} - 1(x-1)(x-1)^{s-1}) \int_0^\infty (p^{s-1}F(p)/\Gamma(s)(1-e^{-p}))\,dp\,ds,$$

$$f(x) = 0 \text{ for } x < A_1.$$

These solutions are easily established on noting that for $\phi(v) = (1-e^{-v})^{-1}$

[27] P. W., p. 37.

[28] Evidently the general formal solution may be written

$$\frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} K(s,x) \int_0^\infty v^{s-1}\phi(v)F(v)\,dv.$$

Theorem 3 applies to the inversion indicated in Eq. 6. 11. In fact it is manifest $|F(v)| < Me^{-A_1 v}/v$ and accordingly $\int_0^\infty |v^{s-1}F(v)(1-e^{-v})|\,dv$ is certainly absolutely convergent for $\mathcal{R}(s) > 2$.

A solution, Eq. 7. 4, formally somewhat similar to that given by Paley and Wiener [29] (who, however, work in the domain of L_2 functions) follows on using the specialization $p = 1$, $\phi(v) = v$ or $\gamma(s, p) = \Gamma(s+1)$ in Eq. 6. This may be rigorously established for $f(x)$ belonging to $L_1(0, \infty)$ as follows:

$$(7.1)\qquad \int_0^\infty f(x)/(1+x)^{s+1}dx = [1/\Gamma(s+1)]\int_0^\infty e^{-\lambda}\lambda^s F(\lambda)\,d\lambda = \psi(s)$$

by Theorem 3 and Eq. 5. 1 for $\mathcal{R}(s) > 0$. Writing $x + 1 = e^z$

$$(7.2)\qquad \psi(s) = \int_0^\infty e^{-sz} f(e^z - 1)\,dz.$$

Clearly,

$$\int_0^\infty |f(e^z-1)|\,dz = \int_0^\infty |f(x)/1+x| < \int_0^\infty |f(x)|\,dx < \infty.$$

Hence Eq. 1. 2 applies and

$$(7.3)\qquad f(e^z-1) = (1/2\pi i)\int_{c-i\infty}^{c+i\infty} e^{zs}\psi(s)\,ds \qquad c > 0,$$

or

$$(7.4)\qquad \begin{aligned} f(x) &= (1/2\pi i)\int_{c-i\infty}^{c+i\infty}(x+1)^s\psi(s)\,ds \\ &= (1/2\pi i)\int_{c-i\infty}^{c+i\infty}\frac{(x+1)^s}{\Gamma(s+1)}\int_0^\infty e^{-\lambda}\lambda^s F(\lambda)\,d\lambda ds. \end{aligned}$$

The result holds when x is interior to a neighborhood of bounded variation for $f(x)$.

Here also a generalization is afforded by replacing convergence by summability $(C, 1)$ or Sommerfeld type, in that the last integral of Eq. 7. 4 is summable to $f(x)$ when $f(x+0) + f(x-0)$ has meaning. This observation hinges essentially on the fact that the summability property of Fourier integrals since $f \subset L_1$ is patently directly extensible to Eq. 1. 21 and hence to Eq. 7. 3 and thus to Eq. 7. 4.

UNIVERSITY OF ILLINOIS,
URBANA, ILL.,

PURDUE UNIVERSITY,
LAFAYETTE, IND.

[29] P. W., p. 37-39. (Here references to D. V. Widder's work may be found as well.) The P. W. solution apart from its implication of L_2 function classes and an apparent integration order change, is essentially transformable to the type of Eq. 7. 4 with the specialization.

NOTE ON FORMAL LOGIC.*

By M. H. Stone.

It has been observed that the theory of Boolean algebras assumes a particularly satisfactory algebraic form when developed in terms of the symmetric difference $a + b$ and the product $a \cdot b$ as fundamental operations: for Boolean algebras are then characterized as rings with unit in which every element is idempotent.[1] The close connection between Boolean algebras and the formal (Aristotelian) logic of propositions therefore suggests that a logistic system built up from corresponding operations would be of some interest, and would have the special advantage of reducing the proofs of most logical theorems to simple and essentially familiar algebraic calculations. In the present note we shall develop such a system, based on results of Leśniewski and Bernstein.[2]

Propositions are to be regarded as abstract entities and denoted by the letters $a, b, c, \cdots$. We postulate three primitive operations on propositions, each of which results in a new proposition; and indicate the propositions resulting from their application by $a + b$, $a \cdot b$, and a' respectively. We may read $a + b$ as "a if and only if b" or "a is equivalent to b"; we may read $a \cdot b$ as "a or b"; and we may read a' as "not a." To indicate that a particular proposition a is to be placed on the list of asserted propositions we write $\vdash a$. As primitive assertions, we postulate that for arbitrary propositions a, b, c (whether given directly or expressed as "polynomials" in terms of the postulated operations and other directly given propositions)

$$(1.1) \qquad \vdash [(a + b) + (c + a)] + [b + c]$$
$$(1.2) \qquad \vdash [a + (b + c)] + [(a + b) + c]$$
$$(2.1) \qquad \vdash [a \cdot (b \cdot c)] + [(a \cdot b) \cdot c]$$
$$(2.2) \qquad \vdash [(a + b) \cdot c] + [(c \cdot a) + (c \cdot b)]$$
$$(2.3) \qquad \vdash (a \cdot a) + a$$
$$(3.1) \qquad \vdash [(a + a') \cdot b] + b.$$

* This note was written while the author was a Fellow of the John Simon Guggenheim Memorial Foundation in residence at the Institute for Advanced Study as a temporary member. Received by the Editors January 7, 1937; revised March 22, 1937.

[1] Stone, *Transactions of the American Mathematical Society*, vol. 40 (1936), pp. 37-111, especially pp. 39-48.

[2] Leśniewski, *Fundamenta Mathematicae*, vol. 14 (1929), pp. 1-81; B. A. Bernstein, *Annals of Mathematics* (2), vol. 37 (1936), pp. 317-325.

In order to bring other propositions upon the list of asserted propositions, we postulate the informal deductive rules

(A) if $\vdash a$ and $\vdash a + b$, then $\vdash b$;

(B) if $\vdash a$, then $\vdash a \cdot b$.

The application of these rules will be indicated by the schemes

$$\text{(A)}\ \frac{\begin{array}{l}\vdash a\\ \vdash a+b\end{array}}{\vdash b} \qquad\qquad \text{(B)}\ \frac{\vdash a}{\vdash a \cdot b}.$$

We introduce a relation $=$ between propositions as follows:

DEFINITION 1. $a = b$ *if* $\vdash a + b$.

We can then introduce two further operations either through the definitions [3]

DEFINITION 2. $\vdash [a \,\&\, b] + [(a + b) + (a \cdot b)]$,

DEFINITION 3. $\vdash [a \rightarrow b] + [b + (a \cdot b)]$,

or through the equivalent definitions

DEFINITION 2′. $a \,\&\, b = (a + b) + (a \cdot b)$,

DEFINITION 3′. $a \rightarrow b = b + (a \cdot b)$.

The proposition $a \,\&\, b$ may be read "a and b," the proposition $a \rightarrow b$ may be read "a implies b." We shall see that the interpretations of the primitive and defined operations are all justified by subsequent results.

We commence our investigation by considering the consequences of (1.1), (1.2), (A) and Definition 1. The system so described is due to Leśniewski.[4] We obtain the following fundamental result:

THEOREM 1. *In terms of the operation $+$ and the relation $=$, taken as an equality-relation, the system under consideration is an additive abelian group in which every element is of order 2. If the zero element of this group*

---

[3] The usual form of definition would be to describe $a \,\&\, b$, $a \rightarrow b$ as abbreviations for $(a + b) + (a \cdot b)$, $b + (a \cdot b)$ respectively. For comments on the present form, which is better suited to our later algebraic considerations, if not to the requirements of a strictly formal logic, see Tarski (Tajtelbaum), *Fundamenta Mathematicae*, vol. 4 (1923), pp. 196-200, especially p. 197.

[4] Leśniewski, *loc. cit.* We write $+$ in place of his $\equiv$.

be denoted by 0, *the statements* $a = 0$ *and* $\vdash a$ *are equivalent. In particular, we have, for all* $a, b, c, d,$

$$\begin{array}{ll} (\alpha) & a = a; \qquad (\alpha') \text{ if } a = b, \text{ then } b = a; \\ (\alpha'') & \text{if } a = b \text{ and } b = c, \text{ then } a = c; \\ (\beta) & \text{if } a = c \text{ and } b = d, \text{ then } a + b = c + d; \\ (\gamma) & a + b = b + a; \qquad (\delta) \; a + (b + c) = (a + b) + c; \\ (\epsilon) & \text{the equation } x + b = a \text{ has } a + b \text{ as a solution.} \end{array}$$

It is well known that the properties (α)–(δ), together with the existence of a solution of the equation $x + b = a$, are characteristic for abelian groups.[5] They imply that the solution of $x + b = a$ is unique (in the sense that $x + b = a$ and $y + b = a$ imply $x = y$) and that the zero element 0 exists and satisfies the equation $x + a = a$ for every a. Thus the property (ϵ) above yields the special relation $a + a = 0$ for every a; in other words, every element is of order 2. Accordingly, we need establish only the properties (α)–(ϵ).

We begin with several lemmas, as follows:

$$\begin{array}{ll} (1.3) & \vdash (a + b) + (b + a); \\ (\text{A}') & \text{if } \vdash a + b, \text{ then } \vdash b + a; \\ (\text{A}'') & \text{if } \vdash b \text{ and } \vdash a + b, \text{ then } \vdash a; \\ (1.4) & \vdash a + a; \\ (1.5) & \vdash [b + a] + [(c + b) + (a + c)]; \\ (1.6) & \vdash a + [b + (b + a)]. \end{array}$$

Ad (1.3). Substituting b, a, b for a, b, c respectively in (1.2), and b, $a + b$, $b + a$ for a, b, c respectively in (1.1), we obtain the scheme

$$(\text{A}) \; \frac{\begin{array}{l} \vdash [b + (a + b)] + [(b + a) + b] \\ \vdash \{[b + (a + b)] + [(b + a) + b]\} + \{(a + b) + (b + a)\} \end{array}}{\vdash (a + b) + (b + a).}$$

Ad (A′). Using (1.3) we have the scheme

$$(\text{A}) \; \frac{\begin{array}{l} \vdash a + b \\ \vdash (a + b) + (b + a) \end{array}}{\vdash b + a.}$$

Ad (A″). We now have the scheme

$$(\text{A}) \; \frac{\vdash b \quad (\text{A}') \, \dfrac{\vdash a + b}{\vdash b + a}}{\vdash a.}$$

[5] See, for instance, van der Waerden, *Moderne Algebra*, Berlin, 1930, vol. I, pp. 15-19.

Ad (1.4). Substituting b, a for a, b respectively in (1.3) and b, a, a for a, b, c respectively in (1.1), we obtain the scheme

$$(\mathrm{A})\ \frac{\begin{array}{l}\vdash (b+a)+(a+b)\\ \vdash [(b+a)+(a+b)]+[a+a]\end{array}}{\vdash a+a.}$$

Ad (1.5). Substituting c, b, a for a, b, c respectively in (1.1), we obtain the scheme

$$(\mathrm{A}')\ \frac{\vdash [(c+b)+(a+c)]+[b+a]}{\vdash [b+a]+[(c+b)+(a+c)]}.$$

Ad (1.6). Using (1.3) and substituting a, b, $b+a$ for a, b, c respectively in (1.2), we obtain the scheme

$$(\mathrm{A}'')\ \frac{\begin{array}{r}\vdash (a+b)+(b+a)\\ \vdash \{a+[b+(b+a)]\}+\{(a+b)+(b+a)\}\end{array}}{\vdash . a+[b+(b+a)].}$$

This completes the proof of the lemmas listed above.

We turn now to the properties (α)–(ϵ), taking them up in a somewhat altered order.

Ad (α). By (1.4) and Def. 1, we have $a=a$.

Ad (α'). By Def. 1, $a=b$ means $\vdash a+b$. Hence $a=b$ implies $\vdash b+a$ by (A′) and thus $b=a$ by Def. 1.

Ad (α''). If $a=b$ and $b=c$, then $\vdash b+a$ and $\vdash c+b$ by (α') and Def. 1. Hence, by using (1.5), we obtain the scheme

$$\begin{array}{l}(\mathrm{A})\ \dfrac{\begin{array}{l}\vdash b+a\\ \vdash [b+a]+[(c+b)+(a+c)]\end{array}}{\vdash (c+b)+(a+c)}\\ (\mathrm{A})\ \dfrac{\vdash c+b}{\vdash a+c.}\end{array}$$

Thus $a=c$ by Def. 1.

Ad (γ). By (1.3) and Def. 1, we have $a+b=b+a$.

Ad (δ). By (1.2) and Def. 1, we have $a+(b+c)=(a+b)+c$.

Ad (β). If $a=c$, we have $\vdash a+c$ by Def. 1. On substituting c, a, b for a, b, c respectively in (1.5), we obtain the scheme

$$(\mathrm{A})\ \frac{\begin{array}{l}\vdash a+c\\ \vdash [a+c]+[(b+a)+(c+b)]\end{array}}{\vdash (b+a)+(c+b).}$$

By Def. 1, we have $b + a = c + b$; and by (γ) and (α'') we infer that $a + b = c + b$. If $b = d$, we can substitute b, c, d for a, b, c, respectively in this equation, obtaining $b + c = d + c$. Applying (γ) and (α''), we infer that $c + b = c + d$. Thus a final application of (α'') yields $a + b = c + d$.

Ad (ϵ). By (α), (β), and (γ), we have $(a + b) + b = (b + a) + b$. By (α'') and (γ), we then have $(a + b) + b = b + (b + a)$. On the other hand, (1.6) and Def. 1 yield $a = b + (b + a)$. Hence (α') and (α'') yield $(a + b) + b = a$.

We still have to establish the equivalence of $a = 0$ and $\vdash a$. From the preceding results, we know that $a = 0$ if and only if $a = a + a$; and that the statements $a = a + a$ and $\vdash a + (a + a)$ are equivalent. Now, using (1.4) and substituting a, a for a, b respectively in (1.6), we obtain the schemes

$$(\mathrm{A}'')\ \frac{\begin{array}{l}\vdash a + a\\ \vdash a + (a + a)\end{array}}{\vdash a} \qquad\qquad (\mathrm{A})\ \frac{\begin{array}{l}\vdash a\\ \vdash a + [a + (a + a)]\end{array}}{\vdash a + (a + a)}.$$

Hence the statements $\vdash a$ and $\vdash a + (a + a)$ are equivalent. It follows that the statements $a = 0$ and $\vdash a$ are equivalent.

We next consider the effect of introducing (2.1), (2.2), (2.3) and (B) into the system studied in Theorem 1. We obtain the following fundamental result:

THEOREM 2. *In terms of the operations* $+$ *and* $\cdot$ *and of the relation* $=$ *of Definition* 1, *the system under consideration is a Boolean ring—that is, a ring (necessarily commutative) in which every element is idempotent.*[6] *In particular, we have, for all* a, b, c,

(η) *if* $a = c$ *and* $b = d$, *then* $a \cdot b = c \cdot d$;

(ζ) $a \cdot b = b \cdot a$; (ι) $a \cdot (b \cdot c) = (a \cdot b) \cdot c$;

(κ) $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$;

(λ) $a \cdot a = a$.

It is well known that the properties (α)–(δ), (η)–(κ), together with the existence of a solution of the equation $x + b = a$ are characteristic for a commutative ring.[7] We may remark that (λ) implies $a + a = 0$ and hence (ϵ): for obvious applications of (α''), (β), (ζ), (κ), and (λ) yield

---

[6] Stone, *loc. cit.*

[7] See, for instance, van der Waerden, *Moderne Algebra*, Berlin, 1930, vol. I, pp. 36-40.

$$\begin{aligned} a+a &= (a+a)\cdot(a+a) = [(a+a)\cdot a]+[(a+a)\cdot a] \\ &= [a\cdot(a+a)]+[a\cdot(a+a)] = [(a\cdot a)+(a\cdot a)]+[(a\cdot a)+(a\cdot a)] \\ &= [a+a]+[a+a]. \end{aligned}$$

We now discuss the indicated properties in a somewhat altered order.

Ad (ι). By (2.1) and Def. 1, we have $a\cdot(b\cdot c)=(a\cdot b)\cdot c$.

Ad (λ). By (2.3) and Def. 1, we have $a\cdot a=a$.

Ad (ζ). By (λ) we have $(a+b)\cdot(a+b)=a+b$. By (2.2) and Def. 1, we have $(a+b)\cdot(a+b)=[(a+b)\cdot a]+[(a+b)\cdot b]$, $(a+b)\cdot a=(a\cdot a)+(a\cdot b)$, $(a+b)\cdot b=(b\cdot a)+(b\cdot b)$. Applying (α'), (α''), (β), (γ), (δ), and (λ) in an obvious manner, we therefore obtain $[(a\cdot b)+(b\cdot a)]+[a+b]=a+b$. Theorem 1 now shows that $(a\cdot b)+(b\cdot a)=0$ and hence that $a\cdot b=b\cdot a$.

Ad (κ). Substituting b, c, a for a, b, c respectively in (2.2) and applying Def. 1, we have $(b+c)\cdot a=(a\cdot b)+(a\cdot c)$. Then by (ζ) and (α''), we have $a\cdot(b+c)=(a\cdot b)+(a\cdot c)$.

Ad (η). If $a=c$, then $\vdash a+c$ by Def. 1. On substituting a, c, b for a, b, c respectively in (2.2), we therefore have the scheme

$$\text{(B)}\ \frac{\vdash\ a+c}{\vdash\ (a+c)\cdot b}$$

$$\text{(A)}\ \frac{\vdash[(a+c)\cdot b]+[(b\cdot a)+(b\cdot c)]}{\vdash(b\cdot a)+(b\cdot c).}$$

By Def. 1, we then have $b\cdot a=b\cdot c$. If $b=d$, we can substitute b, c, d for a, b, c respectively in this equation, obtaining $c\cdot b=c\cdot d$. Applying (α'') and (ζ) in an obvious way, we obtain $a\cdot b=c\cdot d$.

We observe that the informal rule (B) has been used only in the proof of (η). It is therefore of particular interest to note further that (B) can be deduced from (1.1), (1.2), (A), (2.2), and (η), as we shall now show. If $\vdash a$, then $a=0$ by Theorem 1; and $a=a+a$, also by Theorem 1. By (α), (η) and Def. 1, we therefore have $\vdash[a\cdot b]+[(a+a)\cdot b]$. Hence, on substituting $b\cdot a$ for a in (1.4) and a, a, b for a, b, c respectively in (2.2), we obtain the scheme

$$\text{(A}''\text{)}\ \frac{\begin{array}{r}\vdash(b\cdot a)+(b\cdot a)\\ \vdash[(a+a)\cdot b]+[(b\cdot a)+(b\cdot a)]\end{array}}{\vdash\ (a+a)\cdot b}$$

$$\text{(A}''\text{)}\ \frac{\vdash[a\cdot b]+[(a+a)\cdot b]}{\vdash\ a\cdot b.}$$

Thus (η) and (B) may be regarded as equivalent with respect to the primitive propositions and the informal rule (A) of the system.

We now introduce the last of our primitive propositions, (3.1). This proposition is due essentially to Bernstein.[8] We then have

THEOREM 3. *The postulation of* (3.1) *is equivalent to the postulation of a unit e in the Boolean ring of Theorem 2 together with the definition* $a' = a + e$.

By (3.1) and Definition 1, we have $(a + a') \cdot b = b$ for every element b. The element $a + a'$ thus has the properties of a unit in the Boolean ring of Theorem 2. Since two units in a commutative ring are necessarily equal,[9] we see that the unit $a + a'$ is independent of a. Denoting the unit by e, we therefore have $a + a' = e$; and we conclude by Theorem 1 that $a' = a + e$ for every a. On the other hand, if the Boolean ring of Theorem 2 has a unit e and $a' = a + e$, we can apply (α), (α'), (α''), (β), (δ), (ϵ), and (η) to obtain

$$\begin{aligned}(a + a') \cdot b &= [a + (a + e)] \cdot b = [(a + a) + e] \cdot b = [0 + e] \cdot b \\ &= e \cdot b = b;\end{aligned}$$

and Def. 1 then yields $(3.1) \vdash [(a + a') \cdot b] + b$.

The results obtained in Theorems 1, 2, and 3 may now be inverted as follows:

THEOREM 4. *In an additive abelian group, with* 0 *as its zero element, let the truth of the equation $a = 0$ be indicated by $\vdash a$. If this group has the property that every element is of order* 2, *then* (1.1), (1.2), *and* (A) *are theorems; if this group is a Boolean ring under a suitable multiplication, then* (1.1), (1.2), (2.1), (2.2), (2.3), (A), *and* (B) *are theorems; and, if this group is a Boolean ring with unit under a suitable multiplication, then* (1.1), (1.2), (2.1), (2.2), (2.3), (3.1), (A), *and* (B) *are theorems. In each of these cases, $a = b$ if and only if $a + b = 0$ or $\vdash a + b$.*

The proof may be left to the reader.

In order to illustrate the demonstration of logical theorems, in accordance with the principle established in Theorem 1 that the statements $\vdash a$ and $a = 0$ are equivalent, we give the following result:

THEOREM 5. *In the logistic system under consideration we have for all a, b, c*

[8] B. A. Bernstein, *loc. cit.*

[9] See, for instance, van der Waerden, *Moderne Algebra*, vol. I (Berlin, 1930), p. 40.

(4.1) $\vdash (a' \to a) \to a;$

(4.2) $\vdash a \to (a' \to b);$

(4.3) $\vdash [a \to b] \to [(b \to c) \to (a \to c)];$—

together with the informal deductive rule

(C) *if* $\vdash a$ *and* $\vdash a \to b$, *then* $\vdash b$.

Corresponding to the informal rule (C) *we have for all* a, b

(5.1) $\vdash [a \,\&\, (a \to b)] \to b.$

According to Definitions 2, 3 (2′, 3′) and Theorems 1, 2, 3 we have to establish the algebraic identities

(4.1*) $a + [a + (a + e) \cdot a] \cdot a = 0,$

(4.2*) $\big(b + [(a + e) \cdot b]\big) + \big(a \cdot \{b + [(a + e) \cdot b]\}\big) = 0,$

(4.3*) $\big[\{c + (a \cdot c)\} + \{[c + (b \cdot c)] \cdot [c + (a \cdot c)]\}\big]$
$+ \big[\big(b + (a \cdot b)\big) \cdot \big(\{c + (a \cdot c)\} + \{[c + (b \cdot c)] \cdot [c + (a \cdot c)]\}\big)\big] = 0,$

(5.1*) $b + \big[\big(\{a + [b + (a \cdot b)]\} + \{a \cdot [b + (a \cdot b)]\}\big) \cdot b\big] = 0,$

together with the rule

(C*) if $a = 0$ and $b + (a \cdot b) = 0$, then $b = 0$.

Since the operations are to be carried out in a commutative ring with unit e, we can expand these expressions in a familiar way—we can drop all brackets (using the convention that multiplications take precedence over additions), we can write the factors of any product in alphabetical order, and we can drop e as a factor from any product in which it occurs. Our alleged identities then assume the respectively equivalent forms

(4.1**) $a + a \cdot a + a \cdot a \cdot a + a \cdot a = 0,$

(4.2**) $b + a \cdot b + b + a \cdot b + a \cdot a \cdot b + a \cdot b = 0,$

(4.3**) $c + a \cdot c + c \cdot c + a \cdot c \cdot c + b \cdot c \cdot c + a \cdot b \cdot c \cdot c + b \cdot c$
$+ a \cdot b \cdot c + b \cdot c \cdot c + a \cdot b \cdot c \cdot c + b \cdot b \cdot c \cdot c$
$+ a \cdot b \cdot b \cdot c \cdot c + a \cdot b \cdot c + a \cdot a \cdot b \cdot c + a \cdot b \cdot c \cdot c$
$+ a \cdot a \cdot b \cdot c \cdot c + a \cdot b \cdot b \cdot c \cdot c + a \cdot a \cdot b \cdot b \cdot c \cdot c = 0,$

(5.1**) $b + a \cdot b + b \cdot b + a \cdot b \cdot b + a \cdot b \cdot b + a \cdot a \cdot b \cdot b \cdot = 0.$

By application of the special rules $a \cdot a = a$, $a + a = 0$, these relations are seen to be identities, as we wished to prove. As to (C^*), we note that $a = 0$ implies $b + (a \cdot b) = b + (0 \cdot b) = b + 0 = b$, and hence that $a = 0$ and $b + (a \cdot b) = 0$ together imply $b = 0$.

It has been shown by Łukasiewicz and Tarski [10] that a complete logistic system for the Aristotelian logic of propositions can be based on the primitive operations $\rightarrow$ and $'$ with (4.1), (4.2), (4.3) as primitive propositions and (C) as the sole informal deductive rule. Hence the system discussed here contains all of the ordinary logic of propositions. Our system has a similar relation to that of Russell and Whitehead, whose primitive operations correspond to our $\cdot$ and $'$. On the other hand, it can be shown that the Łukasiewicz-Tarski and Russell-Whitehead systems contain ours (under suitable definitions of $+$ and $\cdot$ when not taken as primitive). Since this aspect of the situation is quite familiar, we do not go into detail.

HARVARD UNIVERSITY.

---

[10] Łukasiewicz and Tarski, *Comptes rendus de la Société des Sciences et des Lettres de Varsovie*, Classe III (1930), pp. 30-50; Tarski, *Fundamenta Mathematicae*, vol. 25 (1935), pp. 503-526, especially p. 506.

A CASE OF COLORATION IN THE FOUR COLOR PROBLEM.*

By C. E. Winn.

In studying the four color problem we assume [1] a map divided by a connected trihedral network into a finite number of polygons. Errera [2] has shown that, if no polygon has more than 6 sides, such a map is reducible, i. e. its coloration can be made to depend on that of one or more maps of fewer polygons. In his treatment, however, the reduced figures are not generally maps of the original type, as they may contain polygons of more than 6 sides. Consequently, the resulting map cannot be further reduced by the same method.

In the present paper we shall obtain reductions or sets of reductions which preserve the type of map in the reduced figure, with a view to proving that

I. *A map S containing at most one polygon of more than* 6 *sides can be colored.*

We shall start by obtaining new reductions which, in conjunction with those already known, may be embodied in the result

II. *Any polygon of less than* 7 *sides in an irreducible map must touch a polygon of more than* 6 *sides.*

We now quote the known reductions required here, giving an explanation of how the reduced figures are formed.

A. *A polygon of less than* 5 *sides.*[3]

The reduction is made by removing any side; only, in the case of a

* Received March 31, 1937.

[1] For a general account of the subject see Sainte-Laguë, "Géométrie de situation et jeux," *Mémorial des Sciences Mathématiques*, fasc. 41, and the thesis of M. A. Errera, "Du coloriage des cartes, etc.," Bothy, Ixelles, 1921.

[2] "Une contribution au problème des quatre couleurs," *Bulletin de la Société Mathématique de France*, vol. 53 (1925), p. 42.

[3] A. B. Kempe, "On the geographical problem of the four colors," *American Journal of Mathematics*, vol. 2 (1879), p. 198.

quadrilateral in a 2-ring,[4] we must remove a side bounding the ring, in order to avoid the creation of an isthmus.[5]

It will be seen at once that the reduction of a digon and a triangle diminishes the vertices of the adjacent polygons by two and one respectively, while that of a quadrilateral deprives of one vertex the two polygons abutting the suppressed side.

B. *A ring of 5 polygons or fewer enclosing more than one polygon.*

The reduced maps for a 2- or 3-ring are formed by suppressing the part of the map on one side of the ring.

Two of the reductions for a 4- or 5-ring are made in the same way; and the others are obtained from them by a further removal of two non-adjacent sides of the quadrilateral or pentagon newly formed (i. e. including more than one polygon of the given map).

The reductions A and B are fundamental in as far as they ensure the presence of two successive rings about each polygon of an irreducible map, without which all other known reductions and those given here might fail, should an isthmus occur in the reduced figure.

C. *A polygon completely surrounded by pentagons.*[6]

In the reduction the polygon coalesces with alternate regions on the further side of the ring, except the last two when the number is odd.

D. *A polygon bounded by hexagons and pairs of pentagons, if not by an odd number of hexagons only.*[7]

In the reduction the whole ring is suppressed except the sides joining the *free* vertices (i. e. belonging to *one* polygon only of the ring) of each hexagon and pair of pentagons.

It may be remarked that the reductions of a pentagon or hexagon flanked

[4] A ring may be defined as a cyclic sequence of polygons each of which touches that before and after it, but no other one, in the sequence. The polygons of a 2-ring have 2 separate contacts.

[5] An isthmus occurs at a boundary which can be crossed once only by a closed circuit not meeting the network again.

[6] G. D. Birkhoff, "The reducibility of maps," *American Journal of Mathematics*, vol. 35 (1913), p. 116.

[7] Birkhoff, *loc. cit.* (8) and P. Franklin, "The four color problem," *American Journal of Mathematics*, vol. 44 (1922), p. 225.

by 3 pentagons along adjacent sides are not employed here. In fact the latter might involve more than one polygon of 7 sides or more in the reduced figure.

In order to establish II, we must show how to reduce any ring composed of pentagons or hexagons about a pentagon or hexagon. Actually we need only consider those rings in which the pentagons are *isolated*. For, following a remark of Franklin,[8] *we may derive the reduction of a ring containing a pair or pairs of pentagons from that of a ring obtained on replacing them by pairs of hexagons*, assuming every pair reduced as in D. The derived cases are given in brackets at the head of each reduction.

With a view to saving space our reductions are mainly set forth in tabular form referring to the appropriate figure. Under numbered variants are given all essentially distinct groups of the colors 1, 2, 3, 4 bounding the reduced configuration, i. e. which are neither permutable nor symmetrically equivalent. Obvious abbreviations are employed, such as $b = 1$ to mean that the polygon b bears the color 1.

If the solution for any variant is immediate, we mark the color of the central polygon, from which the ring can be filled in without difficulty. Otherwise one or more similar color chains are assumed (1st column), admitting an alternation of the intermediate complementary chain or else a change due to the absence of former chains (2nd column). There are now two possibilities:

(a) The new distribution may lead to a direct coloring (3rd column), in which case we may suppose the assumed chain absent, and replace the previous change by one affecting a polygon named in the first chain, at the same time allowing for consequent modifications elsewhere (4th column). If a direct solution is not yet available, we may assume another chain in the new distribution, and so on until a coloring is finally obtained.

(b) The change first made may not yield a direct coloring. We then examine one or more chains occurring in the derived scheme, continuing as in (a). The digression is distinguished from the main case by means of brackets, which are closed as soon as a solution is forthcoming. In fact *a completed bracket is tantamount to a direct coloring of the previous scheme.* In complicated cases it may be necessary to insert more than one bracket in succession.

---

[8] *Loc. cit* (8), p. 232.

E. $n5665$ $(n > 5)$. See fig. 1.

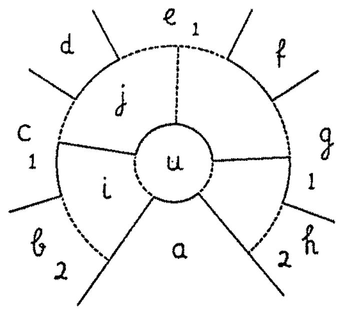


Fig. 1.

Note that, if a were a pentagon, an isthmus would occur in the reduction.

(1) $a = 1$, $d \neq f$ or $d = f = 2$ $u = 2$

(2) $a = 1$, $d = f = 3$

23 b to f	$c = e = 4$	$u = 4$ $u = 3$	$b = 3$; $d, h = 2$ or 3

(3) $a = 3$, $d \neq f$ or $d = f = 4$ $u = 1$

(4) $a = 3$, $d = f = 2$

24 b to h or d	$a = 1$ or $c = 3$	$u = 2$ $u = 1$	$b = 4$, $f = 2$ or 4

(5) $a = 3$, $d = f = 3$

24 b to h	$a = 1$	(1) $u = 1$	$b = 4$

F. 56666 (or 56655 or 56556). See fig. 2.

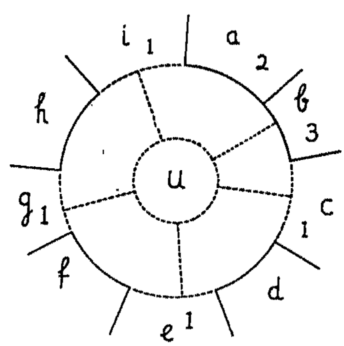


Fig. 2.

(1) $d = 2$ $\qquad$ $u = 1$

(2) $d = 4,\ fh \neq 43$ $\qquad$ $u = 1$

(3) $d = 4,\ f = 4,\ h = 3$

42 d to a	$bc = 13$	$u = 1$	$d = 2$
		$u = 1$	

(4) $d = 3,\ f \neq h$ or $f = h = 4$ $\qquad$ $u = 1$

(5) $d = 3,\ f = h = 3$

32 d to b	$c = 4$	$u = 3$	$d = 2$
		$u = 1$	

(6) $d = 3,\ f = h = 2$

42 f to a	$g = i = 3$		
32 d to b	$c = 4$	$u = 3$	$d = 2$
42 f to d or a	$e = 3,\ i = 1$ or	$u = 3$	$f = 4;\ d, h = 2$ or
		$u = 3$	$f = 2,\ h = 2$ or 4
		$u = 1$	

The next reduction is due to Mr. Choinacki.

G. 66666. See fig. 3.

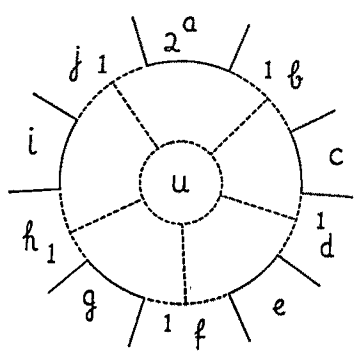


Fig. 3.

(1) cegi not all[9] 2 $\qquad$ $u = 1$

[9] Cp. the reduction of 666666 about a hexagon by Birkhoff, *loc. cit.* [6].

(2) $c = e = g = i = 2$

24 a to e	$b = d = 3$		
[23 i to a	$j = 4$		
{24 e to c or g	$d = 1$ or $f = 3$	$u = 2$	
24 e to a	$b = d = 1$,		
	$c = g = 4$		
(23 a to i	$j = 1$	$u = 1$	$a = 3$, $e = 2$ or 3
		$u = 1$)	$e = 4$
43 j to b or d	$a = 1$, $c = 2$ or 1	$u = 1$	$j = 3$
24 i to g or e	$h = 3$, $f = 1$ or 3	$u = 3$	
24 i to c	$b = j = 1$	$u = 1$	$i = 4$, $a = 2$ or 4
		$u = 3$}	$i = g = 3$ [10]
24 e to a	$d = b = 1$	$u = 1$	
24 e to c	$d = 1$		
(31 b to d or j c	or $a = 4$	$u = 1$	$b = 1$
		$u = 1$)	$e = 4$
		$u = 3$]	$a = 4$; $c, g, i = 2$ or 4
		$u = 1$	

H. 566666 (or 566655 or 565555). See fig. 4.

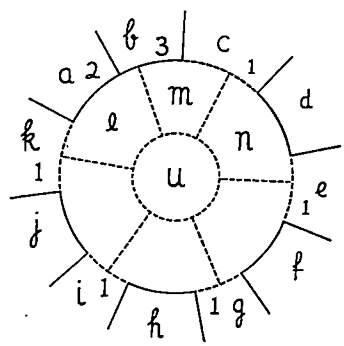


Fig. 4.

(1) If $d = 2$, we get a direct coloring with $u = 1$.

(2) If $d = 4$, we may suppose the 42 chain from d to a present. Hence we have 423 or 342 for lmn, directly or after inverting 31 of bc. One of these yields a solution with $u = 1$, whatever be the colors of f, h, j.

[10] By symmetry the chain 23 from e to g is also absent.

(3) If $d = 3$, the coloring is direct, provided $f = 4$. And, when $f = 2$, the presence of a 24 chain from a to f allows the inversion $bcde = 1313$. Again one of the alternatives affords a solution for any colors of h and j.

(4) If $d = f = 3$, we get a direct coloring for $h = 2$ or 3. When $h = 4$, we proceed as in (2).

The above method can be extended to reduce any polygon enclosed by an *odd* number of hexagons and a pentagon.

The next two reductions are much simplified if we reduce symmetrically by joining the free vertices of the two pentagons across the ring. Unfortunately, however, the two new polygons resulting may both have more than 6 sides, even after A is applied.

J. 565666 (or 565556). See fig. 5.

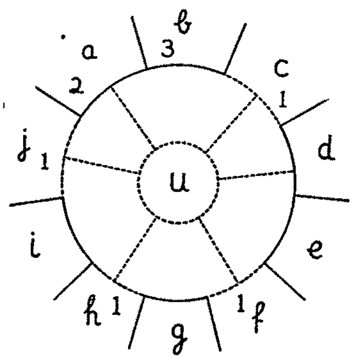


Fig. 5.

(1) $d = 2$, $e = 3$, $gi \neq 34$ $\qquad u = 1$

(2) $d = 2$, $e = 3$, $g = 3$, $i = 4$

12 j to h or f	$i = 3, g = 3$ or 4	$u = 1$	
12 j to c	$b = 4, h = f = 2$	$u = 2$	$ja = 21$
		$u = 1$	

(3) $d = 2$, $e = 4$

43 e to b	$cd = 21$	$u = 1$ or 2	$e = 3$
		(1) or (2)	

(4) $d = 3$, $e = 2$, $gi \neq 34$ $\qquad u = 1$

(5) $d=3,\ e=2,\ g=3,\ i=4$

43 i to b	$ja=21$		
(24 e to i	$fgh=313$	$u=3$	$e=4$
		$u=1)$	$i=3;\ d, g=3$ or 4
		$u=1$	

(6) $d=4,\ e=2,\ g\neq i$ or $g=i=3$ $u=1$

(7) $d=4,\ e=2,\ g=i=2$

23 e to g or b	$f=4$ or $cd=41$	$u=2$	$e=3,\ i=2$ or 3
		$u=1$	

(8) $d=4,\ e=2,\ g=i=4$

43 i to b	$ja=21$		
(42 d to j	$abc=313$	$u=4$	$de=24,\ g=4$ or 2
		$u=1)$	$i=3;\ d, g=4$ or 3
		$u=1$	

(9) $d=4,\ e=3,\ gi\neq 32$ $u=1$

(10) $d=4,\ e=3,\ g=3,\ i=2$

32 e to b	$cd=41$		
(42 c to a	$b=1$	$u=1$	$c=2,\ i=2$ or 4
		$u=1)$	$e=2;\ g, i=2$ or 3
		(6) or (7)	

(11) $d=3,\ e=4,\ g\neq i$ or $g=i=3$

42 e to a	$bcd=131$	$u=1$	$e=2$
		(4) or (5)	

(12) $d=3,\ e=4,\ g=i=2$

42 e to a	$bcd=131$		
(23 a to c	$b=4$	$u=2$	$a=3;\ g, i=2$ or 3
		$u=1)$	$e=2;\ g, i=2$ or 4
		$u=1$	

(13) $d=3,\ e=4,\ g=i=4$

43 i to b	$ja=21$		
(14 a or c to i	$j=3,\ b=2$ or 3	$u=1$	$a=c=4$
43 i to a, g or e	$j=1,\ h=2$ or		
	$h=f=2$	$u=4$	$i=3$
32 d to b or j	$c=1,\ a=4$ or 1	$u=1$	$d=2$
34 i to e	$h=f=2$	$u=4$	$i=4;\ a, g=4$ or 3
		$u=4$)	$i=3$
		(9) or (11)	

K. 566566. See fig. 6.

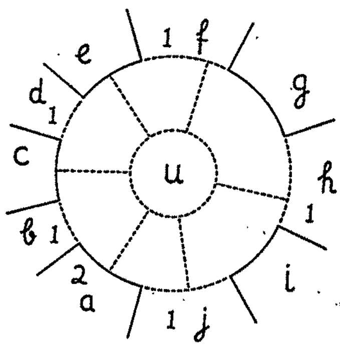


Fig. 6.

(1) $c=e=2$ $u=1$

(2) $c=2,\ e=3,\ gi\neq 42$ $u=1$

(3) $c=2,\ e=3,\ g=4,\ i=2$

32 e to c	$d=4$	$u=2$	$e=2;\ a, i=2$ or 3
		$u=1$	

(4) $c=3,\ e=4,\ g\neq i$ or $g=i=3$ $u=1$

(5) $c = 3,\ e = 4,\ g = i = 2$ ($g = i = 4$ equivalent)

43 e to c	$d = 2$		
[42 e to a	$bc = 31$	$u = 1$	
42 e to i	$f = h = 3$		
{23 a to i	$j = 4$		
(43 j to c, h, f	$ab = 12,\ i = 1,$		
	$g = i = 1$	$u = 3$	$j = 3$
		$u = 3$)	$a = 3,\ c = 3$ or 2
		$u = 3$}	$e = 2,\ g = 2$ or 4
		$u = 2$]	$e = 3$
	equivalent of	(2) or (3)	

(6) $c = 3,\ e = 2,\ g \neq i$ or $g = i = 2$

23 e (or a) to c	d (or b) $= 4$	$u = 2$	$c = 2$
		(1)	

(7) $c = 3,\ e = 2,\ g = i = 3$

24 e to a	$bcd = 313$		
[34 b to d	$c = 2$		
{31 d to f	$e = 4$		
(34 d to b or i	$c = 1$; $a = 2$ or 1	$u = 1$	
34 d to g	$de = 43,\ g = 4$	$u = 2$	$de = 43$
32 e to c	$d = 1$	$u = 3$	$e = 2$; $g, i = 2$ or 3
		$u = 1$)	$d = 1,\ b = 1$ or 3
		$u = 1$}	
34 b to g	$d = 4$	$u = 1$	$b = 4,\ i = 3$ or 4
		$u = 1$]	$e = 4$
		$u = 1$	

(8) $c = 3,\ e = 2,\ g = i = 4$

42 e to a	$bcd = 313$		
{43 b to i (d to g)	$aj = 12$	$u = 1$	
43 b to d	$b = d = 4$		
(41 b to j	$e = 3$	$u = 1$	$bcd = 141$
	equivalent to (7))		$b = 4,\ g = 3$ or 4
		$u = 1$}	$e = 4$; $g, i = 4$ or 2
		(4) or (5)	

L. 565656. See fig. 7.

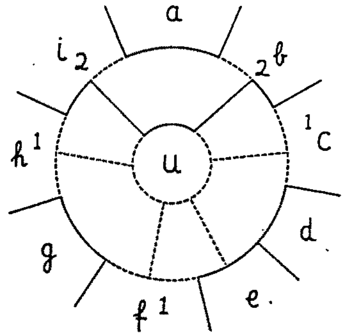


Fig. 7.

(1) $a=1$, $deg \neq 234$ $u=1$

(2) $a=1$, $d=2$, $e=3$, $g=4$

24 d to b or i	$c=3$, $a=1$ or 3	$u=2$ $u=1$	$d=4$, $g=4$ or 2

(3) $a=3$, $d=3$ $u=1$

(4) $a=3$, $de=32$, $g \neq 4$ $u=1$

(5) $a=3$, $de=23$, $g=4$

23 d to b	$c=4$	$u=2$ $u=1$	$d=3$

(6) $a=3$, $de=24$

23 d to b	$c=4$	$u=2$ or 4 $u=1$	$d=3$

(7) $a=3$, $de=42$, $g \neq 3$ $u=2$

(8) $a=3$, $d=4$, $e=2$, $g=3$

34 g to d	$ef=12$	$u=1$ $u=2$	$g=4$, $a=3$ or 4

(9) $a=3$, $de=43$, $g \neq 2$ $u=1$

(10) $a=3$, $de=43$, $g=2$

42 d to b or i	$c=3$, $a=1$ or 3	$u=2$ (4) or (5)	$d=2$, $g=2$ or 4

We proceed to color a map S which may contain one polygon x of more than 6 sides—otherwise we let x be any polygon of S.

First we apply the reduction A to any polygon of less than 5 sides. The removal of any side plainly leaves us with a map of type S; and an exhaustion of the process leads either to a map of 3 digons, which is colorable, or to one free of A polygons.

Next, B is used in conjunction, if necessary, with A. The suppression of the configuration flanking a B ring again leaves a map of type S, as the only new polygon has at most 5 sides. It remains for us to examine the other reductions already explained for a 4 or 5-ring, concerning which we shall justify the following assumptions:

(1) Any 3 polygons abc of a 4-ring R have at least 2 free vertices on either side of R, and therefore at most 4 when R is composed of pentagons and hexagons only.

For, if abc have one vertex only on one side of R, the fourth polygon d has one or more vertices on this side. In the former case R encloses 2 polygons with a total of 8 vertices (unless they form a 2-ring). In the latter case we get a 3-ring with d. Further, if abc have no vertex on one side of R, they touch a quadrilateral or a polygon making a 2-ring with d. Consequently, one of the previous reductions is applicable.

(2) Any 4 polygons of a B 5-ring R have together at least 3 vertices on either side of R, and so 5 at most when R is composed of pentagons and hexagons only.

For, if not, we should get in the same way a ring of 4 or less, or else 2 polygons with a total of 9 vertices or 3 with a total of 13 or 14, according as all or only two of the latter touch the fifth polygon of R.

(3) If x occurs in a B 5-ring at a, the other polygons $bcde$ possess at most 4 free vertices on either side of R, unless both b and e have a free vertex on each side of R.

Otherwise, we can replace b or e by the polygon touching abc or dea. It is easily verified from a figure that the new ring has at most 4 vertices on either side, and on account of (2) surrounds more than one polygon.

Now consider the reduction of a 4-ring composed of $abcd$ which is made by uniting a and c into a new polygon y. If a or $c \equiv x$, there is clearly no gain of vertices except for x. Otherwise, if x still occurs in R, let $b \equiv x$.

We see that $y = 5$ or 6 unless a, c have 3 or 4 vertices outside R (i. e. on the unreduced side), in which case $y = 7$ or 8. In view of (1), however,

d then possesses one or no free vertex outside R, and so becomes a triangle or digon on the further reduction of which we get $y = 6$, as required.

Let us pass to the other reductions of a B ring composed of $abcde$, first supposing $a \equiv x$. As the fusion of a with c or d will yield no gain of vertices elsewhere, we need only deal with the two figures formed by uniting bd (or ce) and be.

In the former case the new polygon y will have $r + 5$ vertices, where b and d have together r free vertices outside R. Thus $y > 6$ only when $r = 2$, 3 or 4.

When $r = 2$, we can further reduce c (< 5) so as to diminish[11] the vertices of y to less than 7.

When $r = 3$, since c, e have, according to (2), at most 2 free vertices outside R, either $c < 5$, $e < 5$ or $c = 2$. In each case their further reduction brings down the number of vertices of y by 2 to less than 7.

When $r = 4$, it follows from (3) that c, e have no free vertices outside R, so that their further reduction diminishes the vertices of y by 3 to less than 7.

The proof for the reduction of a 5-ring in which b and e are united, runs on similar lines, use being made of the further reduction of c and d, when necessary.

Finally, if x does not occur in R, we can no longer appeal to (3). But the above proof holds for $r = 2$ or 3 without change, while the further reduction of a yields the required result when $r = 4$. Moreover, as no polygon of R has more than 6 sides, a single reduction suffices, the other cases being derived by symmetrical or cyclic interchange.

When the reductions A and B are no longer possible, we obtain a map of 3 digons or one of type S for which any of the other reductions quoted or proved are available without the danger of an isthmus appearing in the reduced figure. In each case we shall see that the resulting map is of type S. We apply these reductions in the following order:

(1) D, E, F, G, H, J or K to a polygon u connected with x by the common side of two hexagons. In each reduction the unreduced sides of the two hexagons are separate, so that, except in E, *u and x form a single polygon.* There is consequently no gain of vertices elsewhere.

(2) D, E, F, H, J, K or L to a polygon connected with x by the common

[11] If c has 4 sides left, the previous reductions of 2-rings obviates the creation of an isthmus on the further reduction of c. The same is true of other quadrilaterals to be reduced later.

side of a hexagon and a pentagon. By employing one of the reduced figures or its symmetrical or cyclic equivalent we again ensure that u and x be united in the reduction, except possibly in the case E. Thus in F, if $b \equiv x$, we join the free vertex of the pentagon to those of the other adjacent hexagon. Similarly, a cyclic adjustment of fig. 7 for L may be necessary to make x coalesce with u, the other new polygon having at most 6 vertices.

As regards E, if c, e or $g \equiv x$, then a, being a hexagon, is reduced to a digon. Consequently, the new polygon comprising u, which has not more than 8 vertices, retains at most 6 when the digon is reduced. Again, if b (or h) $\equiv x$ and $c = 5$, the new polygon including c has at most 8 vertices, which we can diminish to 6 or less by further reducing d or f. On the other hand, if $b \equiv x$ and $c = 6$, we apply instead D or H to the hexagon j, which touches c and i in common with x.

(3) C to x, when surrounded by pentagons only. The reduction is seen to produce no gain of vertices except for x.

Since one or other of the above configurations is always present, we have shown how to reduce S to a map of the same type, and hence to obtain the required coloration.

EGYPTIAN UNIVERSITY, CAIRO.

ON THE FUNDAMENTAL GROUP OF A CERTAIN CLASS OF PLANE ALGEBRAIC CURVES.*

By W. S. TURPIN.

1. Introduction. The problem of existence of algebraic functions, z, of two independent variables, x and y, possessing a preassigned branch curve of order n

$$f_n(x, y) = 0 \tag{1}$$

has been considered by Enriques[1] and Zariski.[2] Zariski has shown that, in view of a result of Enriques, this question may be reduced to the consideration of the Poincaré (fundamental) group of the residual space of the branch curve (1) relative to its carrying complex projective plane and the application of the Riemann existence theorem for algebraic functions of one variable having preassigned branch points.

It is sufficient for the theory of algebraic surfaces from the point of view of birational transformations to consider branch curves (1) possessing only ordinary double points and cusps. Zariski has shown that if a curve possesses only ordinary double points then its fundamental group is necessarily cyclic. A simple case of a curve whose fundamental group is not cyclic is that of the branch curve of a cubic surface. If the cubic surface is general, its branch curve is a sextic, f_6, with six cusps on a conic:

$$f_6 : [\phi_3(x, y)]^2 + [\psi_2(x, y)]^3 = 0, \tag{2}$$

where ϕ_3 and ψ_2 are polynomials in x and y of respective degrees 3 and 2. This curve was treated in detail by Zariski and its fundamental group specifically determined.

An obvious generalization of the curve (2) is the curve f_{6m}, of order $6m$, with $6m$ cusps at the intersections of two curves, $\phi_{3m}(x, y) = 0$ and $\psi_{2m}(x, y) = 0$, of orders $3m$ and $2m$ respectively, where m is a positive integer. Such a curve is given by the equation:

$$f_{6m} : [\phi_{3m}(x, y)]^2 + [\psi_{2m}(x, y)]^3 = 0. \tag{3}$$

* Received December 1, 1936.

[1] F. Enriques, "Sulla costruzione delle funzioni algebriche di due variabili possedenti una data curva di diramazione," *Annali di matematica pura ed applicata*, ser. 4, vol. 1 (Nov., 1923), pp. 185-198.

[2] O. Zariski, "On the problem of the existence of algebraic functions," *American Journal of Mathematics*, vol. 51 (1929), pp. 305-328.

The methods of investigation used for the fundamental group of (2) are peculiar to the sextic of this type and do not admit of an extension to the more general class of curves (3). Hence, it was deemed of interest to investigate the structure of the fundamental group of curves of the type (3). One may expect that the methods developed in this investigation may point the way to a possible procedure for other types of curves, for instance for the branch curves of general surfaces of any order in S_3. As it is known, these curves have been completely characterized by B. Segre.[3]

This investigation of the structure of the fundamental group falls under three classifications:

1°. The determination of the fundamental group, $\ddot{G}$, of a degenerate limit curve, $\bar{f}$, of curves (3).

2°. The factorization of the relations of $\bar{G}$ into relations belonging formally to the fundamental group G' of a virtual curve f' with $6m^2$ cusps, of which $\bar{f}$ is a limit curve.

3°. Verification of the fact that $G' = G$, where G denotes the fundamental group of a curve f of type (3).

Our method of attacking the problem in our special case contains the nucleus of a perfectly general procedure, applicable to an arbitrary plane curve with nodes and cusps, provided the complete continuous (irreducible) system $\{f\}$ of curves having the same singularities as f contains some special curve $\bar{f}$, for instance a degenerate curve without multiple components, whose fundamental group can be directly determined. However, to date, a method has not been found for the step 3° of the above procedure that does not appeal to the special geometry of the curves f. This verification is necessary due to the fact that the factorization obtained in 2° is not unique. It seems probable that an equivalence of possible factorizations can be established by purely group theoretic considerations, but, efforts in this direction have not been successful up to the present.

2. General properties of the fundamental group G and its associated group T.[4] Consider the curve f determined by the equation $f(x, y) = 0$ where $f(x, y)$ is a polynomial, of degree n, in the complex variables x and y.

---

[3] B. Segre, "Sulla caratterizzazione delle curve di diramazione nei piani multipli generali," *Mem. Accad. Ital.*, Mat., vol. 1 (1930).

[4] The concepts and results in this section are compiled from the following sources: S. Lefschetz, "Topology," *American Mathematical Society Colloquium Publications*, vol. 12 (1930); E. R. van Kampen, "On the fundamental group of an algebraic curve," *American Journal of Mathematics*, vol. 55 (1933); O. Veblen, "Analysis situs," *American Mathematical Society Colloquium Publications*, vol. 5 (1922); O. Zariski, "On the

We shall be interested in the Poincaré group, G, of the residual space, S, of the curve f relative to its carrying complex projective plane (x, y).

Let us suppose that the coördinate axes have been chosen in such a manner that f does not pass through the point at infinity on the y-axis and that it possesses no multiple components. A generic line of the pencil $\{x = \text{const.}\}$ will thus have n distinct intersections with the curve f. Moreover, lines of this pencil having less than n distinct intersections with f are finite in number. We shall call such lines singular lines of the pencil and denote them by $x = \alpha_i$ $(i = 1, 2, \cdots, \nu)$. Denote by α_0 a point distinct from the set $[\alpha_i]_{i=1,2,\ldots,\nu}$ and let $[\delta'_i]_{i=1,2,\ldots,\nu}$ be a set of non-intersecting loops in the plane of the variable x emanating from $x = \alpha_0$ and surrounding respectively the points of the set $[\alpha_i]_{i=1,2,\ldots,\nu}$. Let $y_k = b_k$ $(k = 1, 2, \cdots, n)$ be the roots of $f(\alpha_0, y) = 0$ and choose loops g_k in the plane of the complex variable y, emanating from the point at infinity and surrounding b_k, in such a manner that g_i and g_j have only the point at infinity in common for $i \neq j$. We denote this base point of the loops g_k by O.

If we define the Poincaré group G of f in the usual manner, it is well known that the loops g_k may be taken as generators of G and that G is independent of the choice of O to within an isomorphism. It is evident that the generators g_k satisfy the following relation:

$$g_1 g_2 \cdots g_n = 1. \tag{4}$$

As x traverses a closed path in its plane, starting from and returning to α_0, the set of points $[b_k]$ will move continuously describing certain paths in the y-plane and returning finally to its original position, although the individual points may have been permuted among themselves. As the points b_k move, their corresponding loops g_k will also vary and this variation is completely determined by the motion of the points b_k and the condition that the set $[g_k]$ should always consist of non-intersecting loops.

If, under cyclical variation of x from α_0, a root y traverses a path from b_i to b_j, the loop g_i is transformed into a loop g'_i surrounding b_j alone and which must therefore be a transform of g_j by some element of G. Moreover, g_i and g'_i are equivalent elements of G and, thus, corresponding to every cyclical variation of x from α_0, we obtain a relation between the elements of G. In particular, if x traverses the loops $[\delta'_i]_{i=1,\ldots,\nu}$ we obtain motions of the roots y_k which yield the relations

problem of existence of algebraic functions," *loc. cit.*; O. Zariski, "On the Poincaré group of rational plane curves." *American Journal of Mathematics*, vol. 58 (1936).

$$(5) \qquad \Phi_{k,i}(g_1, g_2, \cdots, g_n) \equiv g_k^{-1} g_{k,i}^{-1} g_{p_{ki}} g_{ki} = 1 \qquad \begin{pmatrix} k = 1, 2, \cdots, n \\ i = 1, 2, \cdots, \nu \end{pmatrix}$$

where $\{p_{1,i}, p_{2,i}, \cdots, p_{n,i}\}$ is a permutation of $\{1, 2, \cdots, n\}$. It has been shown that the relations (5) together with the relation (4) constitute a complete set of generating relations for the group G.

The class of motions of the set of points $[b_k]$ induced by cyclic variation of x incident to the point a_0 which carry this set into its original position,

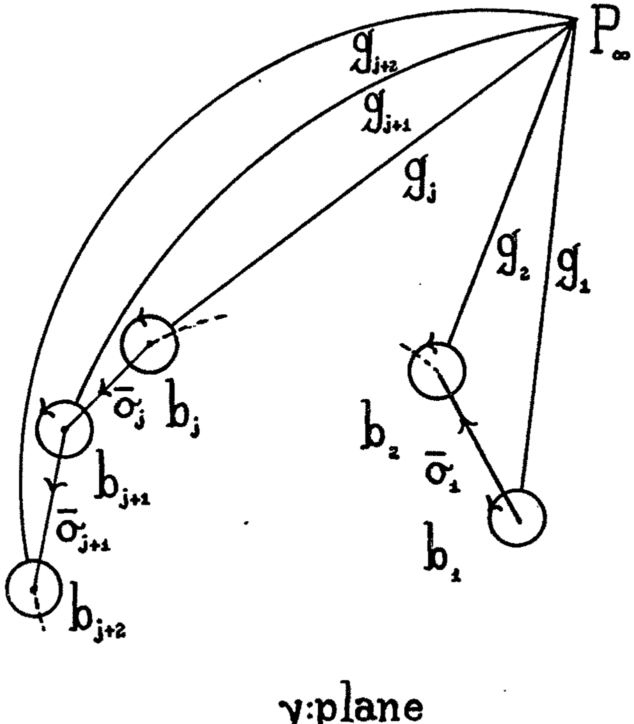


Fig. 1.

in such a way that the roots y_k remain distinct during the motion, constitutes a group of motions T. Two such motions, m_1 and m_2, correspond to the same element of T if the motion m_1 may be deformed into the motion m_2 by suitable deformation of the path of x so that, during the latter deformation, the induced motions keep the roots y_k distinct.

Let $\bar{\sigma}_j$ $(j = 1, 2, \cdots, n-1)$ denote an oriented arc from b_j to b_{j+1} of such a type that the adjacent arcs have only an end point in common and non-adjacent arcs do not intersect (see Fig. 1). Let T_j $(j = 1, 2, \cdots, n-1)$

denote a motion in which the points b_i $(i \neq j, j+1)$ are fixed while the points b_j, b_{j+1} are interchanged, y_j moving from b_j to b_{j+1} along the right-hand edge of $\bar{\sigma}_j$ and y_{j+1} moving from b_{j+1} to b_j along the opposite edge. Any motion, m, belonging to T can be deformed into a motion, m', which may be expressed as a product of the elementary motions T_j. The deformation of m into m' affects not only the paths but also the velocities of the points b_k. The elementary motions T_j satisfy the relations:

$$T_iT_j = T_jT_i, \qquad |i-j| \neq 1 \tag{6}$$

$$T_iT_{i+1}T_i = T_{i+1}T_iT_{i+1}. \tag{7}$$

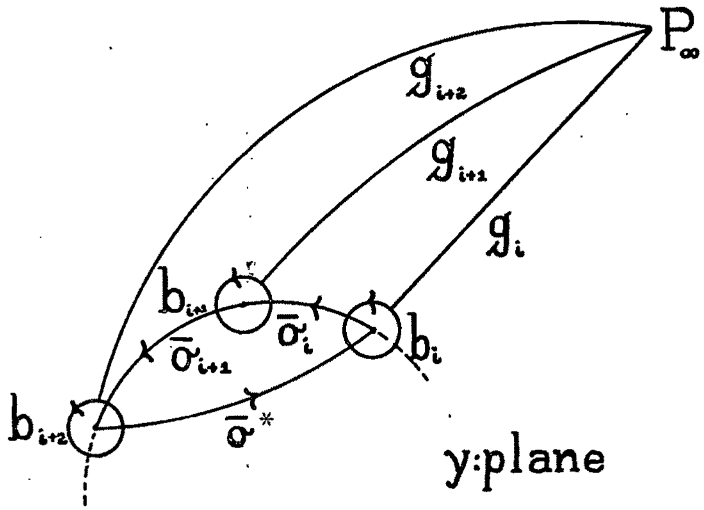


Fig. 2.

The commutative relation, (6), holds due to the fact that, under the assumption $|i-j| \neq 1$, $\bar{\sigma}_i$ and $\bar{\sigma}_j$ have nothing in common and the motions T_i and T_j are thus independent of the order of their performance. The proof of (7) is as follows: Introduce the motion T^* sending y_{i+2} from b_{i+2} to b_i along the right-hand edge of the oriented arc $\bar{\sigma}^*$ and y_i from b_i to b_{i+2} along its opposite edge, where $\bar{\sigma}^*$ is chosen as indicated in Fig. 2. Then the following equalities hold; $T^*T_{i+1} = T_{i+1}T_i = T_iT^*$. Solving these for T^* and equating the solutions, we have that $T_{i+1}T_iT_{i+1}^{-1} = T_i^{-1}T_{i+1}T_i$. Multiplying this relation on the right by T_{i+1} and on the left by T_i, we obtain the desired result.

If we now choose g_1 to be any loop on O surrounding b_1 which does not intersect the arcs $\bar{\sigma}_2, \cdots, \bar{\sigma}_n$, and let g_i be the loop into which g_{i-1} is deformed as y_{i-1} moves from b_{i-1} to b_i along the arc $\bar{\sigma}_{i-1}$ for $(i = 2, \cdots, n)$, then the

motion T_j induces a transformation of the following type on the set of loops $[g_k]$:

$$(8) \qquad t_j : \begin{cases} g'_i = g_i, & (i \neq j, j+1) \\ g'_j = g_{j+1} & \\ g'_{j+1} = g_{j+1}^{-1} g_j g_{j+1} & (j = 1, 2, \cdots, n-1). \end{cases}$$

If we denote the complement of the set of points $[\alpha_i]_{i=1,2,\ldots,\nu}$ relative to the plane of the complex variable x by $C[\alpha_i]$, then, as x describes a cyclic path incident to α_0 in $C[\alpha_i]$, the points $[b_k]$ undergo a motion belonging to the group T, the loops g_k are subjected to the corresponding transformation which, in turn, yields a generating relation of the group G. In particular, the motions $T_{\delta'_i}$ generated as x describes the loops δ'_i induce transformations which yield the relations (5) for G. In consequence, a knowledge of the motions $T_{\delta'_i}$ is sufficient to determine the structure of G.

It will be useful to examine the motion induced on the roots y as x describes a loop about a singular value corresponding respectively to a tangent, ordinary double point and cusp of f.

(A) Suppose $x = \alpha_i$ is a simple tangent to the curve f and that $y_1 = b_1$, $y_2 = b_2$ are the two roots of $f(\alpha_0, y) = 0$ which tend towards coincidence as $x \rightarrow \alpha_i$. Then, as x describes δ'_i, the motion of the roots y_k is equivalent to one in which y_j is fixed for $j > 2$ and the motion of y_1 and y_2 is typified by that of the roots of $y^2 = x$ as x describes the loop $|x| = 1$. This motion has the form T_1. If we do not make the simplifying assumption to the effect that the two roots which approach coincidence as $x \rightarrow \alpha_i$ have consecutive indices, the corresponding motion will have the form $\bar{T}^{-1} T_1 \bar{T}$ where $\bar{T}$ is a product of elementary motions T_j.

(B) Suppose $x = \alpha_i$ is a singular value corresponding to an orrdinary double point of f. Then the motion of the roots induced as x describes δ'_i is equivalent to T_1^2, or, in the general case, to $\bar{T}^{-1} T_1^2 \bar{T}$. Such a singular value α_i may be considered as the limit of two singular values α_i' and α_i'', corresponding to simple tangents of f, at each of which the same pair of roots is permuted.

(C) Suppose $x = \alpha_i$ is a singular value corresponding to a cusp of f. Then the motion of the roots induced as x describes δ'_i is equivalent to T_1^3, or, in the general case, to $\bar{T}^{-1} T_1^3 \bar{T}$. In this case, the singular value α_i can be considered as the limit of three singular values corresponding to simple tangents which have approached coincidence and at all of which the same pair of roots is permuted.

The relations arising between the generators g_i of G due to singular values of types (A), (B) and (C) are, respectively, of the following forms:

(a) $g_1 = g_2$ (b) $g_1 g_2 = g_2 g_1$ (c) $g_1 g_2 g_1 = g_2 g_1 g_2$.

Let us now consider a variable curve f and let $\bar{f}$ denote a limit curve under continuous variation of f. If f and $\bar{f}$ have the same singularities, they are isotopic and, accordingly, possess the same fundamental group. Suppose, however, that as f tends towards $\bar{f}$ it acquires new multiple points or multiple points of higher order. For the sake of simplicity and definiteness, let us suppose that as $f \to \bar{f}$ the simple singular values α_1 and α_2 of f approach coincidence in α so that $\bar{f}$ acquires a new ordinary double point. Any generating relation for f corresponding to a singular value distinct from α_1, α_2 remains a true relation for $\bar{f}$. The relations for f relative to α_1 and α_2, viz. $g_1 = g_2$, are destroyed for $\bar{f}$ since, in the limit when $\alpha_1, \alpha_2 \to \alpha$, both the loops δ'_1 and δ'_2 pass through α. On the other hand, the relation $g_1 g_2 = g_2 g_1$ for f which arises from a circuit, δ', of x about both α_1 and α_2 is not destroyed for $\bar{f}$ since, in the limit, it becomes the relation corresponding to a circuit of x about the singular value α for $\bar{f}$. This reasoning is perfectly general and applies to any new multiple point, or point of higher order, acquired by $\bar{f}$. In fact, the following theorem has been established:

All generating relations of the fundamental group $\bar{G}$ of a limit curve $\bar{f}$ of a variable curve f are also true generating relations for the fundamental group G of f. This theorem holds even if $\bar{f}$ is degenerate provided that it possesses no multiple components.

3. Consideration of the motions, T, occurring at the singular values of a degenerate case of $f_{6m}(x, y)$. We wish to investigate the fundamental group of curves of the type

$$f_{6m}(x, y) \equiv [\phi_{3m}(x, y)]^2 + [\psi_{2m}(x, y)]^3 = 0 \tag{3}$$

where ϕ_{3m} and ψ_{2m} are polynomials in the two complex variables x and y of respective degrees $3m$ and $2m$. It is supposed that the intersections of $\phi_{3m} = 0$ and $\psi_{2m} = 0$ are distinct. If the curves $\phi_{3m} = 0$ and $\psi_{2m} = 0$ are general, the curve $f_{6m} = 0$ will possess $6m^2$ cusps at the intersections of the curves $\phi_{3m} = 0$ and $\psi_{2m} = 0$ and no further singular points. For a general choice of coördinate axes, f_{6m} will possess, under these restrictions, $6m^2$ distinct critical values corresponding to cusps and $6m(3m - 1)$ distinct critical values corresponding to tangents.

We postpone the consideration of (3) temporarily and consider a degenerate case of a curve of this type, namely,

$$(9) \qquad \bar{f}'_{6m}(x, y) \equiv y^{6m} - x^{3m} = 0.$$

The critical values for this function are $x = 0$ and $x = \infty$. The singularity corresponding to each of these values consists of $3m$ branches having simple contact and vertical branch tangent. We now proceed to ascertain the motion

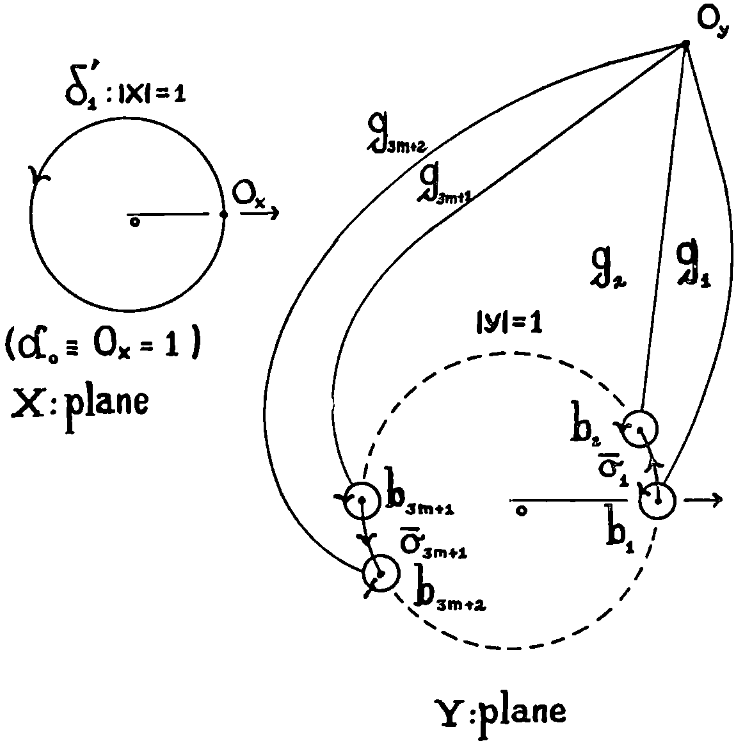


Fig. 3.

T' induced as x describes a loop about the critical value $x = 0$. Let us choose $x = \alpha_0$ to be $x = 1$ and as δ'_1 choose $|x| = 1$. Now, the roots of $f'_{6m}(0, y) = 0$, which we denote by b_k, are the $6m$-th roots of unity. Let us select loops g_i in the y: plane as indicated in Fig. 3. Then as x describes the loop δ'_1, a motion T' is set up which sends b_i into the diametrically opposite point b_{3m+i} along the arc $\overrightarrow{b_i b_{3m+i}}$. Let T_i denote the elementary motion corresponding to the arc $\bar{\sigma}_i$ of Fig. 3. Then the motion T' may be expressed in terms of the elementary motions T_i as follows:

$$T' = (T_{6m-1}T_{6m-2} \cdots T_2T_1)^{3m}. \tag{10}$$

Since the loop δ'_1 may also be considered as a loop about $x = \infty$, the motion of the roots induced as x describes a loop about the latter critical point is again equivalent to T'.

4. Factorization of the motion T'^2 into elementary motions belonging to virtual cusps and simple tangents. The curve (9) is a member of the continuous subsystem of the system $\{\bar{f}_{6m}\}$ consisting of the curves

$$\bar{f}^*_{6m} : y^{6m} = (x - a)^{3m}(x - b)^{3m}, \tag{11}$$

and is isotopic to the general curve of this subsystem. In fact, the equations

$$x' = \frac{x - \tau b}{\tau x + 1 - \tau(1 + a)} \qquad y' = \frac{y}{\tau x + 1 - \tau(1 + a)} \tag{12}$$

where τ is a parameter, define for each value of τ such that

$$\tau b - \tau(1 + a) + 1 \neq 0,$$

a non-degenerate collineation π_τ in the projective (x, y) plane which carries the curve (9) into a curve of the system $\{\bar{f}^*_{6m}\}$. Moreover, π_0 is the identity while $\pi_1(\bar{f}^*_{6m})$ is the curve (9).

The curve $\bar{f}^*_{6m}$ is a limit curve of the general curve, f_{6m}, and we have, in the pencil $\{x = \text{const.}\}$, only two singular lines for $\bar{f}^*_{6m}$, $x = a$ and $x = b$. These singular lines each absorb a certain number of singular lines with respect to the general curve, f_{6m}, and since the singular lines $x = a$ and $x = b$ can be interchanged by a continuous variation of the curve $\bar{f}^*_{6m}$ in the system (11), it follows that each of these lines absorbs $3m^2$ lines $x = \text{const.}$ passing through cusps of f_{6m} and $3m(3m - 1)$ simple tangents of f_{6m}. It must therefore be possible to factor the motion T'^2 into a product of motions $\bar{T}_i$, $6m^2$ of which correspond to cusps and the remaining $6m(3m - 1)$ of which correspond to simple tangents. We proceed to exhibit formally one such factorization of T'^2, making use of the relations (6) and (7), and we shall show afterwards that this factorization actually belongs to the curve $\bar{f}^*_{6m}$ considered as a limiting case of the general curve of the system $\{f_{6m}\}$.

We first write T'^2 in the form

$$T'^2 = (T_{6m-1}T_{6m-2} \cdots T_1)^{6m} = [(T_{6m-1}T_{6m-2} \cdots T_1)^2]^{3m}. \tag{13}$$

It will be useful to establish two lemmas.

LEMMA 1. $T_jT_{j-1}T_{j-2}(T_jT_{j-1})^{-1} = (T_{j-1}T_{j-2})^{-1}T_jT_{j-1}T_{j-2}$.

Proof. Making use of an alternate form of (7) we have

$$T_jT_{j-1}T_{j-2}T_{j-1}^{-1}T_j^{-1} = T_jT_{j-2}^{-1}T_{j-1}T_{j-2}T_j^{-1}$$

and in virtue of (6)

$$T_jT_{j-2}^{-1}T_{j-1}T_{j-2}T_j^{-1} = T_{j-2}^{-1}T_jT_{j-1}T_j^{-1}T_{j-2}.$$

Repeating the first process, we obtain

$$T_{j-2}^{-1}T_jT_{j-1}T_j^{-1}T_{j-2} = T_{j-2}^{-1}T_{j-1}^{-1}T_jT_{j-1}T_{j-2}$$

which is the desired result.

LEMMA 2. $(T_jT_{j-1})^2 = (T_j^{-1}T_{j-1}T_j)T^3{}_{j-1}$.

Proof. Using (7), we may write

$$T_jT_{j-1}T_jT_{j-1} = T_{j-1}T_jT^2{}_{j-1} = (T_{j-1}T_jT_{j-1}^{-1})T^3{}_{j-1}$$

and making use of an alternate form of (7) we are enabled to write the last member in the desired form.

Let us now write $(T_{6m-1}T_{6m-2}\cdots T_1)^2$ in the following manner:

$$\begin{aligned}\{[T_{6m-1}\cdots T_1][(T_{6m-1}T_{6m-2})^{-1}(T_{6m-4}T_{6m-5})^{-1}&\\ \cdots (T_2T_1)^{-1}(T_{6m-1}T_{6m-2})^2(T_{6m-4}T_{6m-5})^2&\\ \cdots (T_2T_1)^2(T_{6m-1}T_{6m-2})^{-1}(T_{6m-4}T_{6m-5})^{-1}&\\ \cdots (T_2T_1)^{-1}][T_{6m-1}\cdots T_1]\}.&\end{aligned}$$

This arrangement is possible due to the fact that the center bracket reduces to the identity in virtue of the relation (6). If we again make use of (6), this expression may be rearranged as follows:

$$\begin{aligned}\{[&T_{6m-1}T_{6m-2}T_{6m-3}\cdot(T_{6m-1}T_{6m-2})^{-1}T_{6m-4}T_{6m-5}T_{6m-6}\cdot(T_{6m-4}T_{6m-5})^{-1}\\ &\cdots T_5T_4T_3\cdot(T_5T_4)^{-1}T_2T_1(T_2T_1)^{-1}][(T_{6m-1}T_{6m-2})^2(T_{6m-4}T_{6m-5})^2\\ &\cdots(T_2T_1)^2][(T_{6m-1}T_{6m-2})^{-1}T_{6m-1}T_{6m-2}(T_{6m-4}T_{6m-5})^{-1}\cdot T_{6m-3}T_{6m-4}T_{6m-5}\\ &\cdots(T_2T_1)^{-1}\cdot T_3T_2T_1]\}.\end{aligned}$$

If we perform the obvious cancellations and apply Lemma 1 to the last bracket, we obtain:

$$\begin{aligned}\{[T_{6m-1}T_{6m-2}T_{6m-3}(T_{6m-1}T_{6m-2})^{-1}\cdots T_5T_4T_3(T_5T_4)^{-1}]\,[(T_{6m-1}T_{6m-2})^2&\\ \cdots(T_2T_1)^2]T_{6m-3}T_{6m-4}T_{6m-5}(T_{6m-3}T_{6m-4})^{-1}\cdots T_3T_2T_1(T_3T_2)^{-1}]\},&\end{aligned}$$

and, applying Lemma 2 to the elements of the middle bracket, we are thus enabled to write T'^2 in the form

$$(14)\quad T'^2 = \{[T_{6m-1}T_{6m-2}T_{6m-3}(T_{6m-1}T_{6m-2})^{-1}\cdots T_5T_4T_3(T_5T_4)^{-1}]$$
$$[(T^{-1}_{6m-1}T_{6m-2}T_{6m-1})T^3{}_{6m-2}(T^{-1}_{6m-4}T_{6m-5}T_{6m-4})T^3{}_{6m-5}\cdots(T_2^{-1}T_1T_2)T_1{}^3]$$
$$[T_{6m-3}T_{6m-4}T_{6m-5}(T_{6m-3}T_{6m-4})^{-1}\cdots T_3T_2T_1(T_3T_2)^{-1}]\}^{3m},$$

which is the desired type of factorization.

5. A function $f^*{}_{6m}$, of type (3), whose critical values correspond to induced motions given by the factors of (14). Consider the curve

$$(15)\qquad F_{6m}(x,y) \equiv (x^{3m}-1)^2-(y^{2m}-1)^3 = 0.$$

The critical values of the function $F_{6m}(x,y)$ are divided into two classes according to the following classification:

1. $x = a_j = \omega^j{}_{3m}$ $(j = 0, 1, \cdots, 3m-1)$, where $\omega^j{}_{3m}$ denote the $3m$-th roots of unity. Each of these critical values corresponds to a line containing $2m$ cusps of the curve $F_{6m} = 0$, each having a vertical cusp tangent.
2. $x = c_{k,l} = |\,2\,|^{1/6m}\cdot e\left[i\pi\dfrac{(-1)^k+8l}{12m}\right]$ for $k = 0, 1$ and $l = 0, 1, \cdots, 3m-1$. Each of these values corresponds to a flex tangent having contact of order $2m-1$ with the curve $F_{6m} = 0$.

Let us choose $x = 0$ as common origin of loops δ'_j, δ'_{kl} in the x: plane selected as indicated in Fig. 4. The roots $y_k = b_k$ of $F_{6m}(0,y) = 0$ are given by:

$$b_{1+3j} = e\left[i\pi\frac{(6j-1)}{6m}\right]$$
$$b_{2+3j} = |\,2\,|^{1/2m}\cdot e^{i\pi j/m}$$
$$b_{3+3j} = e\left[i\pi\frac{(6j+1)}{6m}\right]$$

for $j = 0, 1, \cdots, 2m-1$. We now choose oriented arcs $\bar{\sigma}_k$ as indicated in the diagram of the y: plane of Fig. 4 and choose loops g_k, surrounding b_k, in the manner outlined in Sec. 4. Let us examine the motion induced on the roots y_k as x describes one of the loops δ' in the x: plane. This examination will be somewhat simplified if we also consider the auxiliary curve

$$(16)\qquad Y^3 = X^2$$

obtained from the curve $F_{6m}(x,y) = 0$ by setting $Y = y^{2m}-1$ and $X = x^{3m}-1$. The critical values of the function $F_{6m}(x,y)$ are evidently divided into three classes: the values a_j corresponding to the cusp of (16), the values $c_{0,l}$ corresponding to the ordinary value $X = i$ and the values $c_{1,l}$ corresponding to the

ordinary value $X = -i$. The initial value, $x = 0$, of the loops δ' corresponds to the ordinary value $X = 1$. It is therefore clear that as x describes a loop δ'_j, X describes a loop Δ, and, as x describes a loop δ'_{kl}, X describes a loop Δ_k where Δ, Δ_0, Δ_1 are loops in the X: plane emanating from $X = 1$ and surrounding respectively the values $X = 0, i, -i$. If we denote the roots Y_n of (16) for $X = 1$ by B_n where $B_n = e[(2\pi i/3)(n-2)]$ $(n = 1, 2, 3)$, then, clearly, the image of the points b_{n+3j} in the Y: plane is the point B_n for all values of j. Let us define elementary motions T_i of the points b_k with reference

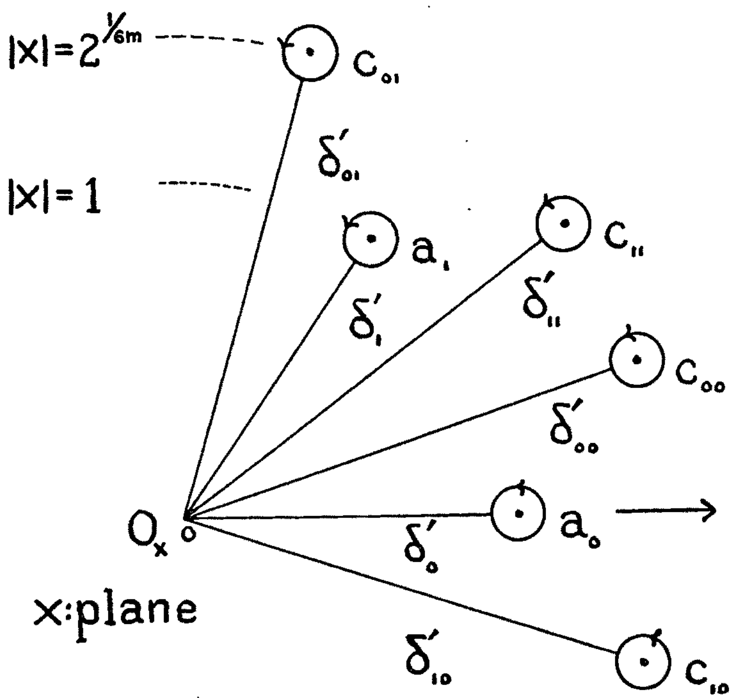


Fig. $4x$.

to the oriented arcs $\bar{\sigma}_k$ of Fig. 4, and elementary motions T^*_i of the points B_k with reference to arcs $\bar{\Sigma}_i$ chosen as positively oriented arcs of the unit circle joining B_i to B_{i+1} for $i = 1, 2$. As x describes the loop δ'_s, X describes the loop Δ and, therefore, the points B_n undergo the motion $(T^*_2T^*_1)^2$. It therefore follows that the image points, b_{n+3j}, in the y: plane are subjected to the motion $(T_{2+3j}T_{1+3j})^2$ for all values of j. Thus, as x describes the loop δ'_s, we have that the corresponding motion of the points b_k is given by

$$(17) \qquad (T_2T_1)^2(T_5T_4)^2 \cdots (T_{6m-1}T_{6m-2})^2$$

for $s = 0, 1, \cdots, 2m-1$.

Suppose x describes the loop δ'_{ks}; then X describes the loop Δ_k. Since Δ_k does not surround a critical value of the function (16), it follows that the motion T^* of the points B_n is the identity. Consequently, the motion of the points b_k must leave the set $[b_{n+3j}]_{(j)}$ invariant for each n. This motion is due to the flex line $x = c_{k,s}$ of the pencil $\{x = \text{const.}\}$ and interchanges the points $b_{(3-2k)+3j}$ in a cyclical manner. This motion has the following description in terms of the elementary motions T_i:

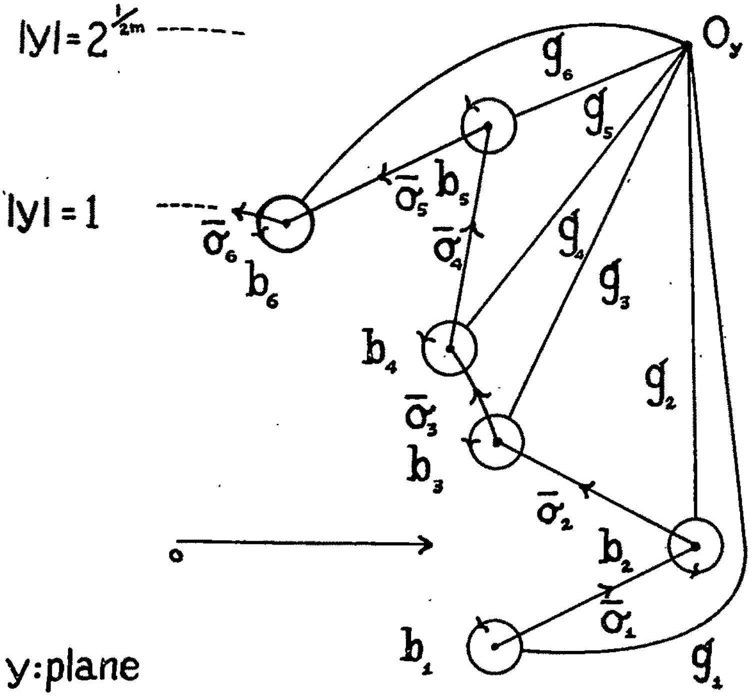


Fig. 4y.

$$\text{(18)} \quad T^{(k)} = \prod_{j=0}^{2m-2}{}^{R} \left(T_{6m+2k-3j-3} T_{6m+2k-3j-4} T_{6m+2k-3j-5} (T_{6m+2k-3j-3} T_{6m+2k-3j-4})^{-1}\right)$$

where $\prod^{R}$ indicates multiplication on the right.

For the curve $F_{6m}(x, y) = 0$, certain of the singular lines of the pencil $\{x = \text{const.}\}$ are coincident. These coincidences are not intrinsic and are due to the special choice of the coördinate axes. In order to eliminate these coincidences from the considerations, we proceed in the following manner.

If we denote by $F_{6m}(\theta)$ the function obtained by rotating F_{6m} through a positive angle θ, then, for small values of θ, the curve $F_{6m}(\theta) = 0$ will have cusps near to the cusps of $F_{6m} = 0$; however, the cusp tangents will no longer

be vertical. In fact, the critical value $x = a_j$ of F_{6m} has associated with it critical values $x = a_{jk}$ corrresponding to cusps of $F_{6m}(\theta) = 0$ and values $x = \bar{a}_{jk}$ corresponding to simple tangents of the curve $F_{6m}(\theta) = 0$, where a_{jk} and $\bar{a}_{jk}$ are in the vicinity of a_j, for $k = 1, 2, \cdots, 2m$.

When a singular line of the pencil $\{x = \text{const.}\}$, which passes through a cusp having this singular line as cusp tangent, breaks up into a pair of singular lines, one of which is a simple tangent and the other a line through the cusp, the motion $(T_{s+1}T_s)^2$ breaks up into either the product of $T^3{}_{s+1}$ and $T^{-1}_{s+1}T_sT_{s+1}$ or of $T_s{}^3$ and $T^{-1}_{s+1}T_sT_{s+1}$, according as which of the two, essentially distinct, possible choices of loops in the x: plane is selected. If we are interested merely in the generating relations of the fundamental group and not in the actual motion of the roots, it is indifferent which of these choices is made, due to the fact that the resulting relations among the generators of the fundamental group are the same in both cases.

The singular lines of the pencil $\{x = \text{const.}\}$ passing through flexes of $F_{6m} = 0$ break up into distinct singular lines passing through points of simple tangency of $F_{6m}(\theta) = 0$. Once more, there is ambiguity concerning the determination of the actual motions corresponding to these singular lines, the motions again being dependent on the choice of loops. However, the flex as a unit imposes the same relations on the generators g_k as does the corresponding number of distinct simple tangents which approach coincidence to form the flex. Consequently, for any choice of loops, the motions corresponding to these simple tangents must impose the same relations on the elements g_k as the factors of the flex motion (18) which correspond to simple tangents. Thus, from the standpoint of relations on the generators g_k, it is sufficient to treat the set of singular lines which approach coincidence in a flex tangent as a unit and merely say that the imposed relations are those of the original flex, the individual motions corresponding to the tangents being left out of the consideration entirely.

To summarize, $F_{6m}(\theta)$ will have the following properties:

1°. Possesses $6m^2$ critical values a_{jk} corresponding to cusps of the curve $F_{6m}(\theta) = 0$ such that, for a proper choice of loops in the x: plane surrounding these values, the induced motions are $(T_{3k+1})^3$ where $k = 0, 1, \cdots, 2m - 1$ and $j = 0, 1, \cdots, 3m - 1$.

2°. Possesses $6m^2$ critical values $\bar{a}_{jk}$ corresponding to simple tangents of the curve $F_{6m}(\theta) = 0$, such that, for a proper choice of loops in the x: plane, the induced motions are $T^{-1}_{3k+2}T_{3k+1}T_{3k+2}$ for $k = 0, 1, \cdots, 2m - 1$ and $j = 0, 1, \cdots, 3m - 1$.

3°. Possesses $6m$ sets of $2m - 1$ critical values c_{klt} corresponding to

simple tangents of the curve $F_{6m}(\theta)=0$ of such a type that, for a proper choice of loops in the x: plane, the induced motions relative to a set c_{klt} for k and l fixed impose the same relations on the generators g_j as do the motions

(19) $$T_{6m+2k-3t-3}T_{6m+2k-3t-4}T_{6m+2k-3t-5}(T_{6m+2k-3t-3}T_{6m+2k-3t-4})^{-1}$$

for $t=0,1,\cdots,2m-2$, where k and l may assume any of the values $k=0,1$; $l=0,1,\cdots,3m-1$.

These properties are exactly those desired for the function f^*_{6m} and, accordingly, we define f^*_{6m} to be the function $F_{6m}(\theta)$.

6. Determination of the structure of the fundamental group G^* of the curve $f^*_{6m}(x,y)=0$. The component transformations of t'^2 associated with the motions corresponding to the critical values of the function $f^*_{6m}(x,y)$ give rise to relations among the generators g_k of G^* of the following types:

1°. From the transformations associated with the motions (19), we obtain the relations

(20. j) $$g_{6m-3j-2}g_{6m-3j-1}g_{6m-3j}=g_{6m-3j-1}g_{6m-3j}g_{6m-3j+1}$$

and the relations

(21. j) $$g_{6m-3j}g_{6m-3j+1}g_{6m-3j+2}=g_{6m-3j+1}g_{6m-3j+2}g_{6m-3j+3}$$

where $j=1,2,\cdots,2m-1$.

2°. The transformations $t^3_{6m-3j+1}$ give the relations

(22. j) $$g_{6m-3j+1}g_{6m-3j+2}g_{6m-3j+1}=g_{6m-3j+2}g_{6m-3j+1}g_{6m-3j+2}$$

for $j=1,2,\cdots,2m$.

3°. The transformations $t^{-1}_{6m-3j+2}t_{6m-3j+1}t_{6m-3j+2}$ yield the relations

(23. j) $$g_{6m-3j+1}=g_{6m-3j+3} \text{ for } j=1,2,\cdots,2m).$$

In addition, the generators g_k also satisfy the trivial relation

(4) $$g_1g_2\cdots g_n=1.$$

If we now apply (23. $j+1$) to (20. j) we have

$$g_{6m-3j-2}g_{6m-3j-1}g_{6m-3j-2}=g_{6m-3j-1}g_{6m-3j-2}g_{6m-3j+1}$$

and, on application of (22. $j+1$) to the left-hand member of this expression, we obtain

$$g_{6m-3j-1}g_{6m-3j-2}g_{6m-3j-1}=g_{6m-3j-1}g_{6m-3j-2}g_{6m-3j+1}$$

whence

(24) $$g_{6m-3j-1}=g_{6m-3j+1} \text{ for } j=1,\cdots,2m-1.$$

In the same way, if we apply (23. j) to (21. j) we obtain

$$g_{6m-3j}g_{6m-3j+1}g_{6m-3j+2}=g_{6m-3j+1}g_{6m-3j+2}g_{6m-3j+1}$$

and, by application of $(22.j)$ to the right-hand member, this expression reduces to

$$g_{6m-3j}g_{6m-3j+1}g_{6m-3j+2} = g_{6m-3j+2}g_{6m-3j+1}g_{6m-3j+2}$$

whence,

$$(25) \qquad g_{6m-3j} = g_{6m-3j+2} \text{ for } j = 1, \cdots, 2m-1.$$

On combining the relations (23), (24) and (25) we obtain

$$(26) \qquad g_1 = g_3 = \cdots = g_{6m-1}; \qquad g_2 = g_4 = \cdots = g_{6m}.$$

If we make use of the relations (26), we are enabled to eliminate the generators $g_3, g_4, \cdots, g_{6m}$ from the relation (4) obtaining

$$(27) \qquad (g_1g_2)^{3m} = 1.$$

Thus, G^* may be generated by two elements g_1 and g_2 satisfying the relations $(22.2m)$ and (27). The relation $(22.2m)$ gives as a consequence the relation $(g_1g_2)^3 = (g_1g_2g_1)^2$ and conversely.

Let us define new elements U and V in the following manner:

$$(28) \qquad U = g_1g_2g_1; \qquad V = g_1g_2.$$

Then the relations:

$$(29) \qquad U^2 = V^3; \qquad U^{2m} = 1,$$

together with the defining relations (28), give as consequences the relations (27) and $(22.2m)$. Hence, it is possible to generate the fundamental group, G^*, of the curve $f^*_{6m}(x, y) = 0$ by the two elements U and V satisfying the relations (29).

7. Conclusion. The curve $f^*_{6m} = 0$ is of the same type as $f_{6m} = 0$. Moreover, the number and kind of singularities are the same for both curves and the restrictive hypothesis to the effect that the cusps of $f_{6m}(x, y) = 0$ should be distinct is also satisfied for $f^*_{6m}(x, y) = 0$. Consequently, it is clear that the curves are isotopic and, therefore, that their fundamental groups possess the same structure. Hence, we may conclude that the fundamental group G of the curves

$$(3) \qquad f_{6m}(x, y) \equiv [\phi_{3m}(x, y)]^2 + [\psi_{2m}(x, y)]^3 = 0$$

may be generated by two elements U and V, of respective orders $2m$ and $3m$, satisfying the relations $U^2V^{-3} = U^{2m} = 1$.

THE JOHNS HOPKINS UNIVERSITY,
BALTIMORE, MARYLAND.

GEOMETRY OF TURBINES, FLAT FIELDS, AND DIFFERENTIAL EQUATIONS.*

By Edward Kasner and John De Cicco.

In this paper, we study the geometry of the oriented lineal elements of a plane. We give additional results to those found in a paper by the senior writer entitled "The group of turns and slides and the geometry of turbines," published in 1911 in the *American Journal of Mathematics*, vol. 33, pp. 193-202. The present paper, however, can be read independently of the earlier paper.

We define ∞^1 elements to be a *series* of elements; this includes a union or curve as a special case. Of course, ∞^2 elements form a *field* of elements, which corresponds to a differential equation of first order $F(x, y, y') = 0$. A *turbine* is the series, which is obtained by converting each element of an oriented circle into one having the same point and a direction making a fixed angle α with the original direction. A *flat field* is the field that is obtained from the totality of all elements which are determined by either the set of all oriented circles containing a given element, or as a special case, the set of all oriented lines, which are parallel to and possess the same orientation as a given line. We desire to study the relationships between general series, general fields (differential equations), turbines and flat fields.

For the analytic representation of an element, it will be found convenient to use two systems of coördinates, called the cartesian and hessian coördinate systems respectively. The cartesian coördinates of an element E are either (x, y, y') or else (x, y, θ), where (x, y) are the cartesian coördinates of the point of the element and θ is the inclination of the line of the element. The hessian coördinates of an element are (u, v, w) where v is the perpendicular from the origin to the line of the element E, u is the angle between the perpendicular and the initial line, and w is the distance between the foot of the perpendicular and the point of the element.

The final theorems constitute wide extensions of Scheffer's[1] theory of isogonal trajectories and equi-tangential trajectories, including his two dual theorems as very special cases.

The *main theorems* in our paper are those numbered 8, 14, 15, 16, 19,

* Received November 13, 1936; Revised December 18, 1936.

[1] *Mathematische Annalen*, vol. 60 (1905), pp. 491-531.

20, 23, 30, 33, 35, 36. The theory of *conjugate differential equations* (possessing the same ∞^2 osculating circles, Theorem 19) is noteworthy.

The turbine. A turbine is the set of oriented elements which are obtained by converting each oriented element of an oriented circle into one having the same point and a direction making a fixed angle α with the original direction. We call the turbine non-linear or linear according as the circle is non-linear or linear.

In cartesian coördinates, the equations of a non-linear turbine are

$$x = a + R \sin(\theta + \alpha), \qquad y = b - R \cos(\theta + \alpha);$$

and in hessian coördinates, the equations are

$$v = a \cos u + b \sin u + r, \qquad w = -a \sin u + b \cos u + s,$$

where

$$r = R \cos \alpha, \qquad s = R \sin \alpha.$$

These equations show that the points of the elements of a turbine form a circle, which we call the *outer circle* of the turbine; and that the lines of the elements are the tangent lines of a circle, which we call the *inner circle*.

In cartesian coördinates, the equations of a linear turbine are

$$x \cos u + y \sin u = v, \qquad \theta + \alpha = u + \pi/2 + 2k\pi,$$

where u and v are constants and, in hessian coördinates, the equations of a linear turbine are

$$U = u - \alpha + 2k\pi, \qquad V \cos \alpha + W \sin \alpha = v.$$

We obtain the following two theorems:

THEOREM 1. *Two elements determine a unique turbine.*

THEOREM 2. *Two turbines possess either no common elements or one common element.*

If a number of elements are all on a turbine, we shall say that these elements are *co-turbinal.*

If a number of turbines all have one element in common, we shall say that these turbines are *co-elemental.*

Let T be the turbine, which is obtained by converting each element of the oriented circle C into one having the same point and a direction making a fixed angle α with the original direction. Then the turbine NT is defined to be the *conjugate turbine* of T, if it is obtained by converting each element of the

oriented circle C into one having the same point and a direction making a fixed angle $-\alpha$ with the original direction. (By means of a certain representation R of the elements of the plane by the points of space, studied in the paper of 1911, conjugacy is defined as polarity with respect to a basic null-system N).

In hessian coördinates, the conjugate turbine NT of the non-linear turbine T

$$v = a\cos u + b\sin u + r, \qquad w = -a\sin u + b\cos u + s,$$

is

$$v = a\cos u + b\sin u + r, \qquad w = -a\sin u + b\cos u - s.$$

The conjugate turbine NT of the linear turbine T

$$U = u - \alpha + 2k\pi, \qquad V\cos\alpha + W\sin\alpha = v,$$

is

$$U = u + \alpha + 2l\pi, \qquad V\cos\alpha - W\sin\alpha = v.$$

THEOREM 3. *The conjugate turbines of two given turbines possess no common elements or one common element according as the two given turbines possess no common elements or one common element.*

The flat field. The totality of elements determined by the set of all oriented circles, which contain a given element, is called a non-linear flat field. The given element is called the center or the central element of the flat field. *All the elements of the field are thus co-circular with a fixed (central) element.*

In cartesian coördinates, the non-linear flat field is given by

$$\theta = -\gamma - 2\arctan\frac{x-\alpha}{y-\beta} + 2k\pi,$$

where (α, β, γ) are the cartesian coördinates of the element which is the negative (or reverse) of the given element contained in the oriented circles. In hessian coördinates the non-linear flat field is given by

$$w = (v + b)\tan\tfrac{1}{2}(a - u) + c,$$

where (a, b, c) are the hessian coördinates of the element, which is the negative of the given element.

The set of all elements obtained by setting $u =$ constant, say α, is called a linear flat field.

It is easy to prove the following theorems:

THEOREM 4. *Three elements determine a unique flat field.*

THEOREM 5. *Two flat fields have in common one and only one turbine.*

THEOREM 6. *Three flat fields have in common one and only one element, or else they have a turbine in common.*

The envelopes of a one parameter family of series of elements. We define ∞^1 elements to be a series of elements. The points of the elements of a series form a curve which we call the point-curve of the series, and the lines of the elements of a series are the tangent lines of a curve which we call the line curve of the series.

Now we consider a one parameter family of series of elements. Let us determine the envelope of the one parameter family of point-curves and the envelope of the one parameter family of line-curves of the one parameter family of series of elements.

Now consider any particular series of the one parameter family of series. The point of intersection of the envelope of the one parameter family of point-curves and of the point-curve of this particular series belongs to an element of this particular series. This element is defined to be an element of the point-envelope of the one parameter family of series of elements.

Again consider any particular series of the one parameter family of series. The common tangent line of the envelope of the one parameter family of line-curves and of the line-curve of this particular series contains an element of this particular series. This element is defined to be an element of the line envelope of the one parameter family of series of elements.

If the one parameter family of series is given in cartesian coördinates by the equations

$$y = f(x, t), \qquad \theta = g(x, t),$$

where t is the parameter, then the point-envelope is given by the equations

$$y = f(x, t), \qquad \theta = g(x, t), \qquad f_t(x, t) = 0.$$

If the one parameter family of series is given in hessian coördinates by the equations

$$v = F(u, t), \qquad w = G(u, t),$$

where t is the parameter, then the line envelope is given by the equations

$$v = F(u, t), \qquad w = G(u, t), \qquad F_t(u, t) = 0.$$

THEOREM 7. *For the point envelope and the line envelope of a one parameter family of series of elements to be identical, it is necessary and sufficient that either the one parameter family of series be a one parameter family of oriented curves; or, if the one parameter family of series is given in cartesian coördinates by the equations*

$$x = \phi(\theta, t), \qquad y = \psi(\theta, t),$$

where t is the parameter, the eliminant with respect to θ of the two equations

$$\phi_t(\theta, t) = 0, \qquad \psi_t(\theta, t) = 0,$$

be identically zero, or if the one parameter family of series is given in hessian coördinates by the equations

$$v = f(u, t), \qquad w = g(u, t),$$

where t is the parameter, the eliminant with respect to u of the two equations

$$f_t(u, t) = 0, \qquad g_t(u, t) = 0,$$

be identically zero.

If a family of series of elements is given in cartesian coördinates by the equations

$$x = \phi(\theta, t), \qquad y = \psi(\theta, t),$$

and, if the eliminant with respect to θ of the two equations

$$\phi_t(\theta, t) = 0, \qquad \psi_t(\theta, t) = 0,$$

is identically zero, we define the series, which is obtained from the equations

$$x = \phi(\theta, t), \qquad y = \psi(\theta, t),$$

where $\theta = \theta(t)$ is the common solution of the two equations

$$\phi_t(\theta, t) = 0, \qquad \psi_t(\theta, t) = 0,$$

to be the envelope of the family of series of elements.

If a one parameter family of series of elements is given in hessian coördinates by the equations

$$v = f(u, t), \qquad w = g(u, t),$$

and if the eliminant with respect to u of the two equations

$$f_t(u, t) = 0, \qquad g_t(u, t) = 0,$$

is identically zero, we define the series, which is obtained from the equations

$$v = f(u, t), \qquad w = g(u, t),$$

where $u = u(t)$ is the common solution of the two equations

$$f_t(u, t) = 0, \qquad g_t(u, t) = 0,$$

to be the envelope of the one parameter family of series of elements.

It is easy to prove that the above two definitions are equivalent.

THEOREM 8. *The necessary and sufficient condition that the one parameter family of turbines*

$$v = a(t)\cos u + b(t)\sin u + r(t), \qquad w = -a(t)\sin u + b(t)\cos u + s(t)$$

possess an envelope is that

$$a'^2 + b'^2 = r'^2 + s'^2.$$

Moreover the envelope is unique and it is given by the equations

$$\cos u = \frac{-a'r' - b's'}{a'^2 + b'^2}, \qquad \sin u = \frac{a's' - b'r'}{a'^2 + b'^2},$$
$$v = a\cos u + b\sin u + r, \qquad w = -a\sin u + b\cos u + s.$$

For the eliminant of the equations

$$a'\cos u + b'\sin u + r' = 0, \qquad -a'\sin u + b'\cos u + s' = 0,$$

is obviously the above condition. The series is unique since $\cos u$, $\sin u$, v, w satisfy linear equations. The theorem follows.

From Theorem 8 we obtain the following:

THEOREM 9. *The one parameter family of conjugate turbines does or does not possess an envelope according as the given one parameter family of turbines does or does not possess an envelope.*

THEOREM 10. *The necessary and sufficient conditions that the one parameter family of turbines*

$$v = a(t)\cos u + b(t)\sin u + r(t), \qquad w = -a(t)\sin u + b(t)\cos u + s(t),$$

be all co-elemental are

$$a'^2 + b'^2 = r'^2 + s'^2, \qquad a'b'' - a''b' = r's'' - r''s'.$$

The tangent turbines of a series of elements. Any series of elements, which has the property that consecutive elements are non-parallel, may be given in hessian coördinates by the equations

$$v = v(u), \qquad w = w(u).$$

In what follows, we use hessian coördinates.

Let u and $u + \Delta u$ determine the two elements of the series

$$(u, v, w) \text{ and } (u + \Delta u, v(u + \Delta u) = v + \Delta v, w(u + \Delta u) = w + \Delta w).$$

Since these two elements cannot be parallel, they determine a unique non-linear turbine, which is given by the parameter values

$$
\begin{aligned}
a &= \frac{1}{2\sin\frac{1}{2}\Delta u}\left[-\Delta v\sin\tfrac{1}{2}(2u+\Delta u)-\Delta w\cos\tfrac{1}{2}(2u+\Delta u)\right],\\
b &= \frac{1}{2\sin\frac{1}{2}\Delta u}\left[\Delta v\cos\tfrac{1}{2}(2u+\Delta u)-\Delta w\sin\tfrac{1}{2}(2u+\Delta u)\right],\\
r &= v+\tfrac{1}{2}\Delta v+\tfrac{1}{2}\Delta w\cot\tfrac{1}{2}\Delta u,\\
s &= w+\tfrac{1}{2}\Delta w-\tfrac{1}{2}\Delta v\cot\tfrac{1}{2}\Delta u.
\end{aligned}
$$

The limiting turbine (of the above set of turbines), as Δu approaches zero, is given by the parameter values

$$
\begin{aligned}
a &= -v'(u)\sin u - w'(u)\cos u,\\
b &= v'(u)\cos u - w'(u)\sin u,\\
r &= v(u) + w'(u),\\
s &= -v'(u) + w(u).
\end{aligned}
$$

We call this turbine the *tangent turbine* of the series at the element (u, v, w).[2]

It is found that the rate of turning with respect to the arc length of the point-curve of the series of the element of the series is $\pm 1/R$ where R is the radius of the outer circle.

THEOREM 11. *The necessary and sufficient conditions that a one parameter family of turbines be a set of tangent turbines of a series of elements, are that the one parameter family of turbines be not a co-elemental family of turbines and possess an envelope. Moreover the envelope is the series to which the turbines are the tangent turbines.*

The proof of Theorem 11 follows immediately from a consideration of the equations for the parameter values of the tangent turbines of a series.

From Theorem 7 and Theorem 11 we obtain

THEOREM 12. *The necessary and sufficient conditions, that a single infinitude of turbines be a set of tangent turbines to a series, are* (1) *that the set of turbines be not co-elemental and,* (2) *if the turbines are not all circles, the line and point-envelopes of the turbines be coincident and, if the turbines are all circles, the envelope of the circles consist of one curve.*

THEOREM 13. *If a one parameter family of turbines is a set of tangent turbines, then the conjugate turbines are either a set of tangent turbines or a set of co-elemental turbines.*

---

[2] If the series is a curve (union), the tangent turbine at the element E becomes the osculating circle of the curve.

The osculating flat fields of series of elements. The flat field, which has three consecutive elements in common with a series at an element E of the series, is called the osculating flat field of the series at E.

Let the flat field

$$w = (v + b) \tan \tfrac{1}{2}(a - u) + c,$$

be the osculating flat field of the series

$$v = v(u), \qquad w = w(u),$$

at the element E. Then if α, β, r, s are the parameter values of the tangent turbine of the series at E, we must have

$$\cos a = \frac{\alpha' r' - \beta' s'}{\alpha'^2 + \beta'^2}, \qquad \sin a = \frac{\alpha' s' + \beta' r'}{\alpha'^2 + \beta'^2},$$
$$b = \alpha \cos a + \beta \sin a - r, \qquad c = -\alpha \sin a + \beta \cos a + s.$$

These equations show that the centers (central elements) of the osculating flat fields of the series are the elements of the envelope of the conjugate turbines.

General fields of elements. The two parameter family of elements, which is given in cartesian coördinates by the equation

$$\theta = g(x, y),$$

or in hessian coördinates by the equation

$$w = f(u, v),$$

where f is of period 2π in u, is called a field of elements.

Let the field be given in hessian coördinates by $w = f(u, v)$. Then, if $v = v(u)$, the functions $v = v(u)$, $w = f(u, v(u))$ define a series of elements, which we call a field series of the field.

Now let a field series contain the element $(u, v, w = f(u, v))$. The parameter values of the tangent turbine of this field series at the element $(u, v, w = f(u, v))$ are given by

$$\begin{aligned} a &= -v'(\sin u + f_v \cos u) - f_u \cos u, \\ b &= v'(\cos u - f_v \sin u) - f_u \sin u, \\ r &= v + f_u + v' f_v, \\ s &= f - v'. \end{aligned}$$

If the field is given in cartesian coördinates by $\theta = g(x, y)$, then the parameter values of the tangent turbine of the field series at the element $(x, y, \theta = g(x, y))$ are

$$a = x - y'/(g_x + y'g_y),$$
$$b = y + 1/(g_x + y'g_y),$$
$$R = \pm\sqrt{1+y'^2}/(g_x + y'g_y),$$
$$\cos(\theta + \alpha) = \pm\frac{1}{\sqrt{1+y'^2}}, \qquad \sin(\theta+\alpha) = \pm\frac{y'}{\sqrt{1+y'^2}}.$$

From these formulae, we easily obtain the following *fundamental theorems on the structure of a field* in the neighborhood of any one of its elements.

THEOREM 14. *Consider any field F and any element E of the field. Then we study the totality of series starting at E and contained in the field. From this totality, select that subset of series whose point-loci pass through the point of E in a given direction and whose line loci thereby necessarily touch the line of E at a fixed point. This subset (although it contains ∞^∞ series) determines a unique tangent turbine.*

THEOREM 15. *By varying the given direction in Theorem 14 we thus obtain ∞^1 turbines. These turbines have their centers on a straight line. This line we shall call the central line relative to the given field F and the given element E.*

From Theorem 15 we obtain

THEOREM 16. *The outer circles of the ∞^1 turbines of Theorem 15 form a pencil in the sense of elementary circle geometry, that is, the circles have two points in common. One of the fixed points is, of course, the point of the element E and the other is a new point which we denote by $\bar{P}$. The inner circles of the turbines form a pencil in the sense of higher circle geometry, that is, the circles have two lines in common. One of the fixed lines is of course the line of the element E and the other is a new line which we denote by $\bar{l}$.*

The central line is given either by the equation

$$x(\cos u - f_v \sin u) + y(\sin u + f_v \cos u) + f_u = 0,$$

or by the equation

$$-g_y(X - x) + g_x(Y - y) = 1.$$

The hessian coördinates of this straight line are determined by

$$\cos(U - u) = \frac{1}{\pm\sqrt{1+f_v^2}}, \qquad \sin(U-u) = \frac{f_v}{\pm\sqrt{1+f_v^2}},$$
$$V = \frac{f_u}{\mp\sqrt{1+f_v^2}}.$$

The tangent flat field. We say that the two fields f and F are tangent to each other at a common element E, if any two field series of the fields f and F respectively, which contain the element E and which have the property that either their line curves or their point curves have a common tangent element at E, are such that their tangent turbines at E are identical.

It is then obvious that two fields $w = f(u, v)$ and $w = F(u, v)$ are tangent at a common element E if

$$f_u = F_u, \qquad f_v = F_v.$$

Similarly the two fields $\theta = g(x, y)$ and $\theta = G(x, y)$ are tangent at a common element E, if

$$g_x = G_x, \qquad g_y = G_y.$$

We call the flat field, which is tangent to a field F at an element E, the tangent flat field of the field F at E.

Let the flat field

$$w = (v + b) \tan \tfrac{1}{2}(a - u) + c,$$

be the tangent flat field of the field

$$w = f(u, v),$$

at the element $E(u, v, w = f(u, v))$. Then we must have

$$\begin{aligned} a &= u + 2 \arctan f_v + 2k\pi, \\ b &= -v - 2f_u/(1 + f_v^2), \\ c &= f(u, v) + 2f_u f_v/(1 + f_v^2). \end{aligned}$$

One parameter family of fields envelope—Characteristics. The equation

$$w = f(u, v, a),$$

defines a one parameter family of fields.

The series, which is a field series of each of two consecutive fields of the family, is called a characteristic of the field. The locus of all the characteristics of the one parameter family of fields is a field and we call it the envelope of the one parameter family of fields. The equations

$$w = f(u, v, a), \qquad f_a(u, v, a) = 0,$$

for each a represent a characteristic of the family and, when we eliminate a from the above two equations, the resulting equation represents the envelope of the family of fields.

It is easily seen that the envelope is tangent to each member of the family of fields at all elements of its characteristics.

The locus of the elements, which are common to consecutive characteristics of a one parameter family of fields, is called the edge of regression. The eliminants, with respect to a, of the equations

$$w = f(u, v, a), \qquad f_a(u, v, a) = 0, \qquad f_{aa}(u, v, a) = 0,$$

give the equations of the edge of regression.

It is obvious that the tangent turbines of the edge of regression and any characteristic at a common element are identical.

Developable fields. The envelope of a one parameter family of flat fields is called a developable field. The characteristics of the one parameter family of flat fields are turbines and these turbines are called the generators of the developable field.

Since each flat field is tangent to the envelope along its characteristic, it follows that the tangent flat field to a developable field is the same at all elements of a generator. The edge of regression of the developable is the series to which the generators are the tangent turbines. Moreover, since consecutive generators are consecutive tangent turbines of the edge of regression, the osculating flat field of the series is that flat field of the family which contains these generators. But this flat field is tangent to the developable. Hence, the osculating flat field at any element of the edge of regression is the tangent flat field to the developable field.

THEOREM 17. *For the field $w = f(u, v)$ to be a developable field, it is necessary and sufficient that*

$$(1 + f_v^2 + 2f_{uv})^2 - 4f_{vv}(f_{uu} + f_u f_v) = 0.$$

For the tangent flat field at the element $(u, v, f(u, v))$ has parameter values

$$a = u + 2 \operatorname{arc\,tan} f_v + 2k\pi, \qquad b = -v - 2f_u/(1 + f_v^2),$$
$$c = f + 2f_u f_v/(1 + f_v^2).$$

The necessary and sufficient condition that $b = b(a)$, $c = c(a)$ is then seen to be the above equation.

Conjugate fields of elements. We define the tangent turbines to any field series of a field to be the tangent turbines of the field. In hessian coördinates the tangent turbines of a field are given by the parameter values

$$\begin{aligned}
a &= -v'(\sin u + f_v \cos u) - f_u \cos u,\\
b &= v'(\cos u - f_v \sin u) - f_u \sin u,\\
r &= v + f_u + v' f_v,\\
s &= f - v'.
\end{aligned}$$

The conjugate turbines of the above turbines are given by the parameter values

$$\begin{aligned}
\bar{a} &= a = -v'(\sin u + f_v \cos u) - f_u \cos u,\\
\bar{b} &= b = v'(\cos u - f_v \sin u) - f_u \sin u,\\
\bar{r} &= r = v + f_u + v' f_v,\\
\bar{s} &= -s = -f + v'.
\end{aligned}$$

Since the tangent turbines of the field $w = f(u, v)$ at the element $(u, v, f(u, v))$ all contain the element $(u, v, f(u, v))$, the conjugate turbines of these turbines must also contain an element and it is unique. We call the element $(\bar{u}, \bar{v}, \bar{w})$ the conjugate of the element (u, v, w). If E is any element of the field $w = f(u, v)$, then we denote the conjugate element by $\bar{E}$. The element $\bar{E}$ is given in hessian coördinates by the equations

$$\begin{gathered}
\cos(\bar{u} - u) = -\frac{1 - f_v^2}{1 + f_v^2}, \qquad \sin(\bar{u} - u) = -\frac{2f_v}{1 + f_v^2},\\
\bar{v} = v + 2f_u/(1 + f_v^2),\\
\bar{w} = -f(u, v) - 2f_u f_v/(1 + f_v^2).
\end{gathered}$$

and in cartesian coördinates, the element $\bar{E}$ is given by the equations,

$$\begin{aligned}
\bar{x} &= x - 2g_y/(g_x^2 + g_y^2),\\
\bar{y} &= y + 2g_x/(g_x^2 + g_y^2),\\
\bar{\theta} &= -g(x, y) + 2 \arctan g_y/g_x + (2k + 1)\pi.
\end{aligned}$$

From these equations, we obtain

THEOREM 18. *The necessary and sufficient condition, that the conjugate elements of a field be the elements of a field, is that the given field be non-developable.*

For, it is obvious that the necessary and sufficient condition, that the set of conjugate elements be at most a one parameter family of elements, is

$$(1 + f_v^2 + 2f_{uv})^2 - 4f_{vv}(f_{uu} + f_u f_v) = 0,$$

which means that the field $w = f(u, v)$ must be a developable field. The theorem follows.

If a field is non-developable, we term the field of conjugate elements the *conjugate field,* a fundamental concept in our theory.

From this follows

THEOREM 19. *Each tangent turbine of the conjugate field is the conjugate turbine of each tangent turbine of the given field. From this it follows that two conjugate families of curves have the same osculating circles.*

For a non-developable field, the equations:

$$\cos(\bar{u}-u) = -\frac{1-f_v^2}{1+f_v^2}, \qquad \sin(\bar{u}-u) = -\frac{2f_v}{1+f_v^2},$$
$$\bar{V} = v + 2f_u/(1+f_v^2),$$

define a line transformation. We call it the conjugate line transformation for the field.

For a non-developable field, the equations

$$X = x - 2g_y/(g_x^2+g_y^2),$$
$$Y = y + 2g_x/(g_x^2+g_y^2),$$

define a point transformation. We call it the conjugate point transformation for the field. The following four results are deduced:

THEOREM 20. *For a line transformation to be a conjugate line transformation of a field, it is necessary and sufficient that the corresponding $\bar{E}$ on $\bar{l}$ of any element E on l be in projective involution with the element E' on $\bar{l}$, which is the tangent element on $\bar{l}$ of the oriented circle which contains the element E and which is tangent to the line $\bar{l}$.*

THEOREM 21. *Let a line transformation be a conjugate line transformation. Then it is the conjugate line transformation of a unique field $w=\phi(u,v)$, which contains a given element (u_0, v_0, w_0). Moreover, any other field, of which it is the conjugate line transformation, is obtained by applying a slide to the elements of the field $w=\phi(u,v)$.*

THEOREM 22. *For a point transformation to be the conjugate point transformation of a field, it is necessary and sufficient that the correspondent $\bar{E}$ on $\bar{P}$ of the element E on P be in projective involution with the element E' on $\bar{P}$ which is the tangent element on $\bar{P}$ of the circle which contains the element E and the point $\bar{P}$.*

THEOREM 23. *Let a point transformation be a conjugate point transformation. Then it is the conjugate point transformation of a unique field $\theta=\psi(x,y)$ which contains a given element (x_0, y_0, θ_0). Moreover, any other field, of which it is the conjugate point transformation, is obtained by applying a turn to the elements of the field $\theta=\psi(x,y)$.*

The tangent turbines of a field. Let us consider the tangent turbines of the field. The parameter values of the turbines are

$$
\begin{aligned}
a &= -w(\sin u + f_v \cos u) - f_u \cos u,\\
b &= w(\cos u - f_v \sin u) - f_u \sin u,\\
r &= v + f_u + w f_v,\\
s &= f - w,
\end{aligned}
$$

where the turbine determined by u, v, w is the tangent turbine of any field series which contains the element $(u, v, f(u, v))$ and whose line curve contains the tangent element (u, v, w) at the element $(u, v, f(u, v))$.

By means of the above equations we are able to prove the following theorems:

THEOREM 24. *For the set of tangent turbines of a field to be a three parameter family of turbines, it is necessary and sufficient that the field be not a flat field.*

THEOREM 25. *For the set of tangent turbines of a field to be a two parameter family of turbines, it is necessary and sufficient that the field be a flat field.*

THEOREM 26. *For a two parameter family of turbines to be the tangent turbines of a flat field, it is necessary and sufficient that the conjugate turbines all contain a given element. Moreover, the given element is the center of the flat field.*

THEOREM 27. *The necessary and sufficient condition, that every one parameter family of turbines of the tangent turbines of a field possess an envelope, is that the field be a flat field.*

THEOREM 28. *For the tangent turbines of a field to be field series of the field, it is necessary and sufficient that the field be a flat field.*

Let us now consider the three parameter family of turbines

$$
\begin{aligned}
v &= a(\lambda, \mu, \nu)\cos u + b(\lambda, \mu, \nu)\sin u + r(\lambda, \mu, \nu),\\
w &= -a(\lambda, \mu, \nu)\sin u + b(\lambda, \mu, \nu)\cos u + s(\lambda, \mu, \nu).
\end{aligned}
$$

Since the above set of turbines is a three parameter family of turbines, at least two of the jacobians

$$
\frac{D(a, b, r)}{D(\lambda, \mu, \nu)}, \quad \frac{D(a, b, s)}{D(\lambda, \mu, \nu)}, \quad \frac{D(a, r, s)}{D(\lambda, \mu, \nu)}, \quad \frac{D(b, r, s)}{D(\lambda, \mu, \nu)},
$$

are not identically zero.

A three parameter family of turbines, whose inner circles are all distinct, is called a general three parameter family of turbines. It is seen that the necessary and sufficient condition for a three parameter family of turbines to be a general set of turbines is that the jacobian

$$\frac{D(a, b, r)}{D(\lambda, \mu, \nu)}$$

be not identically zero. Hence, any general three parameter family turbines may be given by the equations

$$v = a \cos u + b \sin u + r, \qquad w = - a \sin u + b \cos u + s(a, b, r).$$

A three parameter family of turbines, such that it consists of turbines which contain an element of a fixed series, is called a co-serial set of turbines.

THEOREM 29. *The necessary and sufficient conditions for a three parameter family of turbines to be a co-serial set of turbines are that the family be a general set of turbines and that the equations*

$$s_a^2 + s_b^2 = 1 + s_r^2,$$
$$(1 + s_r^2)(s_{aa} + s_{bb}) + s_{rr} = (2s_a + s_b s_r) s_{br} + (-2s_b + s_a s_r) s_{ar},$$

be identically satisfied.

The fixed series is uniquely determined and is given by the equation

$$\cos u = \frac{-s_b + s_a s_r}{1 + s_r^2}, \qquad \sin u = \frac{s_a + s_b s_r}{1 + s_r^2},$$
$$v = a \cos u + b \sin u + r, \qquad w = - a \sin u + b \cos u + s(a, b, r).$$

THEOREM 30. *For a three parameter family of turbines to be a set of tangent turbines of a field, it is necessary and sufficient that the family be not a co-serial set of turbines, that the family be a general set of turbines and finally that the equation*

$$s_a^2 + s_b^2 = 1 + s_r^2$$

be identically satisfied.

The field, to which the turbines are the tangent turbines, is unique and it is given by the equations

$$\cos u = \frac{-s_b + s_a s_r}{1 + s_r^2}, \qquad \sin u = \frac{s_a + s_b s_r}{1 + s_r^2},$$
$$v = a \cos u + b \sin u + r, \qquad w = - a \sin u + b \cos u + s(a, b, r).$$

THEOREM 31. *The necessary and sufficient conditions, that a three parameter family of turbines be a set of tangent turbines of a field, are* (1) *that the set of turbines be not a co-serial set of turbines,* (2) *that the set be a general set of turbines, and* (3) *that, with each turbine T of the family and its conjugate turbine NT, there be associated two unique elements E and $\bar{E}$ respectively, where the element E is on T and the element $\bar{E}$ is on NT, such that every one parameter set of enveloping turbines Ω of the family has the property that, either the series, to which the turbines are the enveloping turbines, consists of the elements E, each of which is on a turbine T of Ω or that the series, (to which the one parameter family of enveloping turbines $N\Omega$, each of which is the conjugate turbine of a turbine of Ω, is the enveloping set of turbines) consists of the elements $\bar{E}$, each of which is on a turbine of $N\Omega$.*

THEOREM 32. *If the series, to which the single infinitude of enveloping turbines $N\Omega$ (or Ω) are the enveloping turbines, consists of the element $\bar{E}$ (or E), then every element of the series, to which the single infinitude of enveloping turbines Ω (or $N\Omega$) are the enveloping turbines, is the element of a turbine T (or NT) of the enveloping turbines Ω (or $N\Omega$), which is on the line q, such that the tangent line of the curve of centers of the turbines at the center of T (or NT) is the bisector of the angle, whose sides are the oriented line of $\bar{E}$ (or E) on NT (or T) and the line q.*

A characteristic property of whirl transformations. We begin by considering certain simple operations or transformations on the oriented lineal elements of the plane. A turn T_a converts each element into one having the same point and a direction making a fixed angle a with the original direction. By a slide S_k the line of the element remains the same and the point moves along the line a fixed distance k. *These transformations together generate a continuous group of three parameters, which we call the group of whirl transformations and which we denote by G_3.* It is easily seen that any whirl transformation may be put in the form [3]

$$T_a S_k T_\beta.$$

The slide S_k is

$$\bar{u} = u, \qquad \bar{v} = v, \qquad \bar{w} = w + k.$$

The turn T_a is

$$\bar{u} = u + \alpha, \qquad \bar{v} = v \cos \alpha + w \sin \alpha, \qquad \bar{w} = -v \sin \alpha + w \cos \alpha.$$

[3] See Kasner, *American Journal of Mathematics*, 1911. The name *whirl* for *TST* was suggested by D. Sole in my seminar. Recently this theory has been extended to spherical geometry by K. Strubecker, *Jahr. d. Math. Ver.*, vol. 44 (1934), pp. 184-198, who suggests the term *turbine-rotation* for whirl.

It is then seen that any whirl transformation may be given by the equations

$$\bar{u} = u + \alpha + \beta, \qquad \bar{v} = v\cos(\alpha+\beta) + w\sin(\alpha+\beta) + k\sin\beta,$$
$$\bar{w} = -v\sin(\alpha+\beta) + w\cos(\alpha+\beta) + k\cos\beta,$$

It is found that the only contact transformations of the set of whirl transformations are

$$T_{\pi/2}S_kT_{-\pi/2}; \qquad T_{\pi/2}S_kT_{\pi/2}.$$

The first represents a dilatation D_k; and the second, which may be written D_kT_π, represents a dilatation accompanied by reversal of orientation.

Our group of whirl transformations may be written in the simple form

$$S_kD_dT_\alpha;$$

and hence any whirl transformation is given in the form

$$\bar{u} = u + \alpha,$$
$$\bar{v} = v\cos\alpha + w\sin\alpha + d,$$
$$\bar{w} = -v\sin\alpha + w\cos\alpha + k.$$

We give now, without proof, a new characteristic property of whirl transformation in terms of the central lines of a field.

THEOREM 33. *For an element transformation to be such, that the central lines of every field $w = f(u, v)$ are identical with the central lines of the corresponding field $\bar{w} = \bar{f}(\bar{u}, \bar{v})$, it is necessary and sufficient that the element transformation be a whirl transformation.*

We remark that the group of whirls is isomorphic to the group of motions. These two groups are commutative and together generate a *new group of six parameters*, of considerable interest in the geometry of elements.

Extension of Scheffer's theory of isogonal and equi-tangential trajectories. First we state the following:

THEOREM 34. *If two fields are related by a whirl transformation, then the two conjugate fields are related by a whirl transformation.*

Let F and G be two fields such that G is obtained from F by applying a whirl transformation W to F. Then, by means of the above theorem, we know that there exists a whirl transformation $\bar{W}$ such that the two fields $\bar{F}$ and $\bar{G}$, the conjugate fields of F and G respectively, have the property that

$\bar{G}$ is obtained from $\bar{F}$ by applying $\bar{W}$ to $\bar{F}$. We call $\bar{W}$ the conjugate whirl transformation of W.

Let E_0 and S_0 be a fixed element and a fixed series respectively. There exists a unique one parameter family of transformations T, which is a subset of the group of whirl transformations W, such that any transformation of T carries E_0 into an element of S_0. It follows that with any element E of the plane there is associated a unique series S, such that any transformation of T carries the element E into an element of S. We define S to be the quasi-path series of E for the set of transformations T. It is seen that the set of quasi-path series for the set of transformations T is at most a three parameter family of series. We denote the totality of quasi-path series for the set of transformations T by Σ.

We say that the one parameter family of transformations $\bar{T}$ is the conjugate set of transformations of the set of transformations T, if each transformation of $\bar{T}$ is the conjugate whirl transformation of a transformation of T and conversely. We denote any quasi-path series of $\bar{T}$ by $\bar{S}$ and the totality of quasi-path series of $\bar{T}$ by $\bar{\Sigma}$. We shall say that two series are conjugate with respect to T or $\bar{T}$, if one is a series of Σ and the other is a series of $\bar{\Sigma}$.

Now let us apply T to a one parameter family of curves F_1. We then obtain ∞^1 new one parameter families of curves, or collectively, a two parameter family of curves F_2. Similarly let us apply $\bar{T}$ to the conjugate family of curves $\bar{F}_1$ of the family of curves F_1. From Theorems 33 and 34 we obtain:

THEOREM 35. *Consider any one of the quasi-path series S of the set Σ connected with T. Each element of S determines a curve of the doubly infinite system of curves F_2 generated by applying T to any simply infinite system F_1. The locus of the centers of the ∞^1 circles osculating these curves at these elements is a straight line. Hence these circles touch a certain series $\bar{S}$ of the set $\bar{\Sigma}$ conjugate to Σ with respect to T.*

THEOREM 36. *According to the previous theorem, the system F_2 obtained by applying T to F_1 induces a definite correspondence between the set of series Σ and the conjugate set $\bar{\Sigma}$. There exists another system $\bar{F}_2$, obtained by applying $\bar{T}$ to $\bar{F}_1$, for which this correspondence is precisely reversed.*

If we place on T the restriction that it be a group of transformations, then the quasi-path series become path series and moreover the path series are turbines all congruent to each other. The set of transformations $\bar{T}$ is also a group of transformations and its path series are turbines, which are the conjugate turbines of the turbines which are the path series of T. Thus we

obtain the following two theorems due to Kasner[4] which are themselves generalizations of Scheffer's[5] fundamental theorems on isogonal and equitangential trajectories:

THEOREM 37. *Consider any one of the path turbines S of the set Σ connected with the one parameter group of transformations T. Each element of S determines a curve of the doubly-infinite system F_2 generated by applying T to any simply-infinite system F_1. The locus of the centers of the ∞^1 circles osculating these curves at these elements is a straight line. These circles touch a certain turbine $\bar{S}$ of the set $\bar{\Sigma}$ conjugate to Σ.*

THEOREM 38. *According to the previous theorem, the system F_2 obtained by applying T to F_1 induces a definite correspondence between the set of turbines Σ and the conjugate set $\bar{\Sigma}$. There exists another system $\bar{F}_2$, obtained by applying $\bar{T}$ to $\bar{F}_1$, for which this correspondence is precisely reversed.*

In Theorems 37 and 38, if we let T be first the group of turns and then the group of slides, we obtain Scheffer's theorems on isogonal and equitangential trajectories. (Part of Scheffer's first theorem was discovered by Çesaro). The most general families whose central loci are straight lines have been studied by Kasner[6] (velocity families and the dual type). In this connection we may obtain characterizations of both the conformal and the equilong groups.

COLUMBIA UNIVERSITY,
NEW YORK.

---

[4] *American Journal of Mathematics*, 1911.

[5] *Mathematische Annalen*, 1905.

[6] *Princeton Colloquium*, 1913, 1934, *American Journal of Mathematics*, 1910, and several abstracts in *Bulletin of the American Mathematical Society*, 1930-1935.

PARALLELISM AND EQUIDISTANCE OF CONGRUENCES OF CURVES OF ORTHOGONAL ENNUPLES.*

By R. M. Peters.

It is the purpose of this paper to develop some theorems relating to angular and distantial spreads [1] of congruences of curves of orthogonal ennuples lying in an n-dimensional Riemannian space V_n. We shall be particularly concerned with congruences belonging to a sheaf; that is, to a totality of ∞^{n-1} congruences of which each two cut under a constant angle. It is assumed that the linear element of the V_n is defined by the positive definite quadratic form $ds^2 = g_{ij}dx^i dx^j$, the x's being coördinates of the V_n, and the g_{ij}'s real analytic functions of the x's. The curves considered are assumed to be real and analytic.

We begin by considering the angular spreads (or associate curvatures, in Bianchi's terminology) of n mutually orthogonal congruences C_h, $(h = 1, 2, 3, \cdots, n)$, with respect to any fixed congruence of curves C. Let $\lambda_h|^i$ and ξ^i be respectively the unit vectors tangent to the curves C_h and C. Then

$$\xi^i = \sum_h \cos \alpha_h \lambda_h|^i, \qquad \sum_h \cos^2 \alpha_h = 1,$$

where α_h is the angle between the curves C_h and C. We denote by $\mu_l|^j$ the angular spread vector, and by $1/r_l$ its length, the angular spread, of the curves C_l with respect to the curves C. Then

$$(1) \qquad \mu_l|^j = \lambda_l|^j{}_{,i}\, \xi^i = \sum_h \cos \alpha_h\, \lambda_l|^j{}_{,i}\, \lambda_h|^i = \sum_h \cos \alpha_h\, \mu_{lh}|^j,$$

where $\mu_{lh}|^j$ is the angular spread vector of the curves C_l with respect to the curves C_h, and $\mu_{ll}|^j$ the first curvature vector of the curves C_l. From (1) we conclude

Theorem 1. *If the curves C_l of one congruence are geodesics and are parallel with respect to the curves of every other congruence of the ennuple, then the curves C_l are parallel with respect to the curves of every other congruence in V_n; that is, their tangent vectors form a field of parallel unit vectors.*

---

* Received February 20, 1937.

[1] For the definition and significance of these terms see Graustein, "The geometry of Riemannian spaces," *Transactions of the American Mathematical Society*, vol. 36, no. 3, p. 555, and Peters, "Parallelism and equidistance in Riemannian geometry," *American Journal of Mathematics*, vol. 57 (1935), pp. 103-111.

THEOREM 2. *If the curves of each congruence of the ennuple are geodesics and are parallel with respect to the curves of every other congruence of the ennuple, then their tangent vectors form n fields of parallel vectors.*[2]

Introducing the coefficients of rotation γ_{pqr} of the orthogonal ennuple, we have

$$\mu_{hk}|^i = \lambda_k|^j \lambda_h|^i{}_{,j} = \sum_r \gamma_{hrk} \lambda_r|^i = -\sum_r \gamma_{rhk} \lambda_r|^i, \tag{2}$$

and

$$1/r^2_{hk} = \sum_r \gamma^2{}_{rhk}, \tag{3}$$

where $1/r_{hk}$ is the angular spread of the curves C_h with respect to the curves C_k. Substituting (2) in (1),

$$\mu_l|^j = -\sum_{r,h} \cos\alpha_h \gamma_{rlh} \lambda_r|^j, \tag{4}$$

and

$$1/r_l{}^2 = \sum_{r,h,k} \cos\alpha_h \cos\alpha_k \gamma_{rlh} \gamma_{rlk}. \tag{5}$$

Now let us consider a second orthogonal ennuple of mutually orthogonal congruences of curves $\bar{C}_h$, $(h = 1, 2, \cdots, n)$, with unit tangent vectors $\bar{\lambda}_h|^i$, and let both ennuples belong to a sheaf. Expressing the vectors $\bar{\lambda}_h|^i$ in terms of the vectors $\lambda_k|^i$, we have

$$\bar{\lambda}_h|^i = \sum_k \cos\beta_{hk} \lambda_k|^i,$$

where β_{hk} is the constant angle between the curves $\bar{C}_h$ and C_k. Since the two ennuples are orthogonal, the angles β_{hk} satisfy the relation

$$\sum_h \cos\beta_{hk} \cos\beta_{hl} = \delta_l{}^k, \tag{6}$$

$\delta_l{}^k$ being the Kronecker delta.

Let $\bar{\mu}_h|^j$ be the angular spread vector and $1/\bar{r}_h$ its length, the angular spread, of the curves $\bar{C}_h$ with respect to the curves C. We inquire what relations exist between the two sets of n vectors $\bar{\mu}_h|^j$ and $\mu_h|^j$, and between the quantities $1/\bar{r}_h$ and $1/r_h$:

$$\bar{\mu}_h|^j = \bar{\lambda}_h|^j{}_{,i}\, \xi^i = \sum_k \cos\beta_{hk} \lambda_k|^j{}_{,i}\, \xi^i = \sum_k \cos\beta_{hk} \mu_k|^j. \tag{7}$$

Since $\bar{C}_h$ may be any congruence of the sheaf, we can state

[2] The ennuple is then Cartesian and the V_n is Euclidean. See Graustein, *loc. cit.*, p. 564.

THEOREM 3. *The angular spread vector of any congruence of curves $\bar{C}$ belonging to a sheaf with respect to an arbitrary congruence of curves C is linearly expressible in terms of the angular spread vectors of any orthogonal ennuple of curves C_h of the sheaf with respect to the same curves C, the coefficients of combination being the cosines of the constant angles between the curves C_h and $\bar{C}$.*

THEOREM 4. *If the curves of any orthogonal ennuple of congruences of the sheaf are parallel with respect to an arbitrary congruence of curves C, then the curves of every congruence of the sheaf are parallel with respect to the curves C.*

This result combined with Theorem 2 gives

THEOREM 5. *If the curves of each congruence of any orthogonal ennuple of a sheaf are geodesics, and are parallel with respect to the curves of every other congruence of the ennuple, then every congruence of curves of the sheaf is parallel with respect to every congruence of curves in the V_n.*

For the angular spread $1/\bar{r}_h$ we have

$$1/\bar{r}_h^2 = \sum_{k,l} \cos\beta_{hk} \cos\beta_{hl}\, g_{ij}\, \mu_k|^i\, \mu_l|^j.$$

Summing over h and using (6)

$$\sum_h 1/\bar{r}_h^2 = \sum_h 1/r_h^2. \tag{8}$$

In this result is contained a theorem given by Bortolotti:[3] the sum of the squares of the angular spreads of the curves of an orthogonal ennuple of congruences of a sheaf with respect to an arbitrary congruence of curves is the same for every orthogonal ennuple of the sheaf.

We shall show that corresponding facts hold for distantial spread vectors, providing the curves C, previously chosen arbitrarily, are required to belong to the sheaf.

Let $\nu_l|^j$ be the angular spread vector of the curves C with respect to the curves C_l. Then

$$\begin{aligned} \nu_l|^j &= \lambda_l|^i\, \xi^j{}_{,i} \\ &= \lambda_l|^i \sum_h (\cos\alpha_h\, \lambda_h|^j{}_{,i} + \lambda_h|^j\, \partial\cos\alpha_h/\partial x^i) \\ &= \sum_h (\cos\alpha_h\, \mu_{hl}|^j + \lambda_h|^j\, \partial\cos\alpha_h/\partial s^l), \end{aligned}$$

[3] Bortolotti, "Stelle di congruenze e parallelismo assoluto," *Rendiconti dei Lincei* (6), vol. 9 (1929), pp. 530-538.

where s^l is the arc of the curves C_l and $\partial/\partial s^l$ denotes directional differentiation in the positive direction of the curves C_l.

If we denote by $b_l|^j$ the distantial spread vector [4] of the curves C_l and C,

$$\begin{aligned} b_l|^j &= \mu_l|^j - \nu_l|^j \\ &= \sum_h \left[\cos\alpha_h(\mu_{lh}|^j - \mu_{hl}|^j) - \lambda_h|^j\, \partial\cos\alpha_h/\partial s^l\right] \\ &= \sum_h \left(\cos\alpha_h\, b_{lh}|^j - \lambda_h|^j\, \partial\cos\alpha_h/\partial s^l\right), \end{aligned} \tag{9}$$

where $b_{lh}|^j$ is the distantial spread vector of the congruences of curves C_l and C_h. Formula (9) holds when the curves C are the curves of any congruence in V_n.

We recall that the distantial spread vector of two congruences vanishes identically if and only if the two congruences lie in a family of two-dimensional surfaces, and the curves of each congruence are equidistant with respect to the curves of the other.[5] Hence we conclude from (9)

THEOREM 6. *If the distantial spread vectors formed for the curves C_l and every other congruence of curves of the ennuple are null vectors, then the curves C_l are equidistant with respect to the congruences of curves C which intersect the curves of all congruences of the ennuple at angles which are constant along the curves C_l. In particular, the curves C_l are equidistant with respect to the curves of all congruences belonging to the sheaf.*

If we require that the curves C belong to the sheaf, (9) becomes analogous in form to (1).

$$b_l|^j = \sum_h \cos\alpha_h\, b_{lh}|^j. \tag{10}$$

An ennuple is called a Tchebycheff ennuple [6] if the distantial spread vector formed for each two congruences of the ennuple is a null vector. Hence we have from (10)

THEOREM 7. *If the orthogonal ennuple is an ennuple of Tchebycheff, then the curves of the ennuple are equidistant with respect to every other congruence of curves of the sheaf, and vice versa.*

Returning to the second orthogonal ennuple of curves $\bar{C}_h$ of the sheaf, let $\bar{\nu}_l|^j$ denote the angular spread vector of the curves C with respect to the curves $\bar{C}_l$, where the curves C are again arbitrary.

[4] For the definition of the distantial spread vector, see Graustein, *loc. cit.*, p. 555.

taken in the order named, then

[5] Graustein, *loc. cit.*, p. 559.

[6] Graustein, *loc. cit.*, p. 563.

$$(11) \qquad \bar{\nu}_l|^j = \bar{\lambda}_l|^i \xi^j{}_{,i} = \sum_k \cos \beta_{lk} \lambda_k|^i \xi^j{}_{,i} = \sum_k \cos \beta_{lk} \nu_k|^j.$$

We note that (11) holds regardless of whether or not the angles β_{lk} are constant. That is, (11) gives the relation between the angular spread vectors of an arbitrary congruence of curves C with respect to the curves of two arbitrary orthogonal ennuples.

For the angular spread, $1/\bar{\rho}_l$, of the curves C with respect to the curves $\bar{C}_l$, we have

$$(12) \qquad \begin{aligned} 1/\bar{\rho}_l{}^2 &= \sum_{h,k} \cos \beta_{lh} \cos \beta_{lk} \, g_{ij} \, \nu_h|^i \nu_k|^j \\ &= \sum_{h,k} \cos \beta_{lh} \cos \beta_{lk} \cos \theta_{hk}/\rho_h \rho_k, \end{aligned}$$

where θ_{hk} is the angle between the vectors $\nu_h|^i$ and $\nu_k|^j$, and $1/\bar{\rho}_h$ is the angular spread of the curves C with respect to the curves C_h. In particular, if we take the curves $\bar{C}_l$ as coincident with the curves C, we obtain

$$(13) \qquad 1/\rho^2 = \sum_{h,k} \cos \alpha_h \cos \alpha_k \cos \theta_{hk}/\rho_h \rho_k,$$

$1/\rho$ being the first curvature of the curves C.[7]

From (11) and (13) it follows that if the curves C of an arbitrary congruence are parallel with respect to the curves of any orthogonal ennuple of congruences, then the curves C are parallel with respect to the curves of every congruence in the V_n; that is, the tangents to the curves C form a field of unit parallel vectors. The curves C are then geodesics.

Summing over l in (12) we obtain

$$(14) \qquad \sum_l 1/\bar{\rho}_l{}^2 = \sum_l 1/\rho_l{}^2.$$

This formula gives a theorem corresponding to the one quoted from Bortolotti, the rôles of the curves C and C_h being interchanged.

THEOREM 8. *The sum of the squares of the angular spreads of the curves of an arbitrary congruence with respect to the curves of an orthogonal ennuple is independent of the ennuple chosen.*

Returning to the consideration of distantial spreads we have from (7) and (11) for the distantial spread vector $\bar{b}_l|^j$ of the congruences $\bar{C}_l$ and C, in the order named,

---

[7] This is a generalization of a form of Liouville's formula for geodesic curvature in a V_2 given by Graustein, *Transactions of the American Mathematical Society*, vol. 34, no. 3, p. 571.

$$(15) \qquad \bar{b}_l|^j = \bar{\mu}_l|^j - \bar{\nu}_l|^j = \sum_k \cos\beta_{lk}\, b_k|^j,$$

where the curves C_l and $\bar{C}_k$ now belong to orthogonal ennuples of the same sheaf so that the angles β_{lk} are constants. This result is entirely analogous to formula (7) for the angular spread of the curves $\bar{C}_l$ with respect to the curves C, in both cases the curves C being entirely arbitrary. We draw conclusions analogous to Theorems 3, 4, and 5.

THEOREM 9. *The distantial spread vector of any congruence of curves $\bar{C}$ of a sheaf and an arbitrary congruence of curves C is linearly expressible in terms of the distantial spread vectors of any orthogonal ennuple of curves C_h of the sheaf and the same curves C, the coefficients of combination being the cosines of the constant angles between the curves C_h and $\bar{C}$.*

THEOREM 10. *If, for an arbitrary congruence of curves C and each congruence of any orthogonal ennuple of the sheaf, the distantial spread vector is a null vector, then the distantial spread vector of C and each congruence of the sheaf is also a null vector.*

Furthermore, using Theorem 7 and requiring that the congruence of curves C belong to the sheaf, we have

THEOREM 11. *If an orthogonal ennuple of the sheaf is an ennuple of Tchebycheff, then every congruence of curves of the sheaf is equidistant with respect to every other congruence of the sheaf.*

We note that this result includes that of Theorem 7.

Let $1/b_l$ and $1/\bar{b}_l$ denote respectively the lengths of the distantial spread vectors $b_l|^j$ and $\bar{b}_l|^j$. Multiplying (15) by $g_{ij}\,\bar{b}_l|^i$, summing over l, and using (6) we obtain

$$(16) \qquad \sum_l 1/\bar{b}_l^2 = \sum_l 1/b_l^2,$$

a result analogous to (8) and (14) for angular spreads, which gives the theorem corresponding to Bortolotti's theorem and Theorem 8.

THEOREM 12. *The sum of the squares of the lengths of the distantial spread vectors formed for the curves of an arbitrary congruence of a sheaf and the curves of every congruence of an orthogonal ennuple of the sheaf is independent of the ennuple chosen.*

Instead of the single congruence of curves C, let us now consider n mutually orthogonal congruences C^*_h, not necessarily belonging to the sheaf,

with unit tangent vectors $\xi_h|^i$. In the previous work we attach an h to each symbol formed with respect to the curves C; for example, α_{kh} now denotes the angle between the curves C_k and C^*_h. Formula (5) becomes

$$(17) \qquad 1/r^{*2}_{hk} = \sum_{r,p,q} \cos\alpha_{pk} \cos\alpha_{qk}\, \gamma_{rhp}\, \gamma_{rhq},$$

where we have replaced $1/r_h$, the angular spread of the curves C_h with respect to the curves C, by $1/r^*_{hk}$, the angular spread of the curves C_h with respect to the curves C^*_k. Summing over k, we have

$$(18) \qquad \sum_k 1/r^{*2}_{hk} = \sum_{r,p} \gamma^2{}_{rhp},$$

since the angles α_{pk} satisfy the relation

$$\sum_k \cos\alpha_{pk} \cos\alpha_{qk} = \delta_q{}^p.$$

Using (3), (18) becomes

$$(19) \qquad \sum_k 1/r^{*2}_{hk} = \sum_k 1/r^2{}_{hk},$$

and, summing over h,

$$(20) \qquad \sum_{h,k} 1/r^{*2}_{hk} = \sum_{h,k} 1/r^2{}_{hk}.$$

Incidentally we note that (19) is essentially identical with (14).

Since we have seen from formula (8) that $\sum_h 1/r^{*2}_{hk}$ is independent of the ennuple of curves C_h, and since (20) shows that $\sum_{h,k} 1/r^{*2}_{hk}$ is independent of the ennuple of curves C^*_k, we conclude

THEOREM 13. *The sum of the squares of the angular spreads of each congruence of curves of an orthogonal ennuple of a sheaf with respect to the curves of each congruence of an arbitrary orthogonal ennuple is the same for any choice of both ennuples.*

Formula (18) furnishes incidentally a proof of the known fact that the quantity $\sum_{h,k} 1/r^2{}_{hk}$ is the same for every orthogonal ennuple of the sheaf.[8]

If we denote by $b^*_{hk}|^j$ the distantial spread vector of the curves C_h with respect to the curves C^*_k, (10) becomes

$$(21) \qquad b^*_{hk}|^j = \sum_l \cos\alpha_{lk}\, b_{hl}|^j,$$

where, we recall, the curves C^*_k now belong to the sheaf. Multiplying by $g_{ij}\, b^*_{hk}|^i$ and summing over i, j, k, and h, we have

[8] Graustein, *Transactions of the American Mathematical Society*, vol. 36, no. 3, p. 579, and Bortolotti, *loc. cit.*

$$\sum_{h,k} 1/b^{*2}_{hk} = \sum_{h,k} 1/b^2{}_{hk}, \tag{22}$$

where $1/b^*{}_{hk}$ is the length of the vector $b^*{}_{hk}|^j$.

By (16) $\sum_h 1/b^{*2}_{hk}$ is independent of the ennuple of curves C_h chosen from the sheaf, and by (22) $\sum_{h,k} 1/b^{*2}_{hk}$ is independent of the choice of the ennuple of curves $C^*{}_k$. Hence we have the following

THEOREM 14. *The sum of the squares of the lengths of all the distantial spread vectors of the curves C_h of an orthogonal ennuple of a sheaf formed with respect to all the curves of a second orthogonal ennuple of curves $C^*{}_k$ of the sheaf is independent of the choice of both ennuples.*

Here also we have an indirect proof of the fact that $\sum_{h,k} 1/b^2{}_{hk}$ is the same for every orthogonal ennuple of the sheaf.[9]

Let us now consider the special case when the sheaf contains an orthogonal ennuple of normal congruences. If we take these as the congruences C_h, (18) becomes, since $\gamma_{rhp} = 0$ for r, h, p all distinct,

$$\sum_k 1/r^{*2}_{hk} = \sum_r \gamma^2{}_{rhr} + \sum_r \gamma^2{}_{rhh} = \sum_r \gamma^2{}_{hrr} + 1/r^2{}_{hh},$$

$1/r_{hh}$ being the first curvature of the curves C_h. Summing over h

$$\sum_{h,k} 1/r^{*2}_{hk} = 2 \sum_h 1/r^2{}_{hh}. \tag{23}$$

Now (2) becomes

$$\mu_{hk}|^i = \gamma_{hkk}\, \lambda_k|^i$$

and hence

$$b_{hk}|^i = \gamma_{hkk}\, \lambda_k|^i - \gamma_{khh}\, \lambda_h|^i,$$

$$1/b^2{}_{hk} = \gamma^2{}_{hkk} + \gamma^2{}_{khh}.$$

Summing over h and k,

$$\sum_{h,k} 1/b^2{}_{hk} = 2 \sum_h 1/r^2{}_{hh}.$$

Hence from (20), (22), and (23)

$$\sum_{h,k} 1/b^{*2}_{hk} = \sum_{h,k} 1/b^2{}_{hk} = 2 \sum_h 1/r^2{}_{hh} = \sum_{h,k} 1/r^2{}_{hk} = \sum_{h,k} 1/r^{*2}_{hk}. \tag{24}$$

Equations (24) show that we have to deal with the following five properties:

[9] Graustein, *loc. cit.*, and Bortolotti, *loc. cit.*

(A) $1/r_{hh} = 0$, $(h = 1, 2, \cdots, n)$: Curves C_h geodesics.

(B) $1/r_{hk} = 0$, $(h, k = 1, 2, \cdots, n;\ h \neq k)$: Curves C_h parallel with respect to the curves C_k.

(C) $1/b_{hk} = 0$, $(h, k = 1, 2, \cdots, n)$: Curves C_h equidistant with respect to the curves C_k.

(D) $1/r^*_{hk} = 0$, $(h, k = 1, 2, \cdots, n)$: Curves C_h parallel with respect to the curves C^*_k of any orthogonal ennuple in V_n.

(E) $1/b^*_{hk} = 0$, $(h, k = 1, 2, \cdots, n)$: Curves C_h equidistant with respect to the curves C^*_k of any orthogonal ennuple of the sheaf, and vice versa, and each pair of congruences C_h and C^*_k lying in a family of two-dimensional surfaces.

From (24) we conclude

THEOREM 15. *If an orthogonal ennuple of normal congruences of curves C_h has any one of the above properties, then it also has the remaining four, and all congruences of the sheaf consist of geodesics.*[10]

Let us add a few properties to the above list so as to summarize and extend some of our previous results. We rewrite (D) and (E) in slightly different form:

(D) Curves C_h, $(h = 1, 2, \cdots, n)$, parallel with respect to the curves of every congruence in V_n.

(E) Curves C_h, $(h = 1, 2, \cdots, n)$, equidistant with respect to the curves of every congruence in the sheaf, and vice versa.

(F) The curves of every congruence of the sheaf parallel with respect to the curves of every other congruence in V_n.

(G) The curves of every congruence of the sheaf equidistant with respect to the curves of every other congruence of the sheaf, and each pair of congruences lying in a family of two-dimensional surfaces.

(H) Curves C_h, $(h = 1, 2, \cdots, n)$, all normal.

By Theorem 2, properties (A) and (B) lead to (D); by Theorem 5, these same properties lead to (F), a result superseding the former since (F) includes (D).

By Theorem 7, (C) leads to (E); and by Theorem 11, (C) leads to (G), (G) including (E).

[10] Part of this theorem, to the effect that if (C) holds, then (A) and (B) hold, and conversely, has been proved by Graustein, *loc. cit.*, p. 564, for a non-orthogonal ennuple in which the curves of each two congruences intersect at an angle constant along the curves of both congruences.

We recall that if an orthogonal ennuple is an ennuple of Tchebycheff, it consists of normal congruences.[11] Furthermore, if (B) is valid, then so is (C). Hence we have finally

THEOREM 16. *If either property* (B) *or* (C) *holds, then all the remaining properties hold.*

We shall now denote by p the sheaf containing the curves C_h, and consider a second sheaf p' and the relations between the angular and distantial spread vectors of the curves of the two sheaves with respect to an arbitrary congruence of curves C with unit tangent vectors ξ^i. Let C'_h, $(h = 1, 2, \cdots, n)$, be the curves of any orthogonal ennuple of p', and $\lambda'_h|^i$ their unit tangent vectors. Let β_{hk} be the angle between the curves C'_h and C_k, where β_{hk} is now a variable. Then

$$\lambda'_h|^i = \sum_k \cos\beta_{hk}\, \lambda_k|^i.$$

Denote by $\mu'_h|^i$ the angular spread vector of the curves C'_h with respect to the curves C, and by $\nu'_h|^i$ the angular spread vector of the curves C with respect to the curves C'_h. Then

$$\begin{aligned}\mu'_h|^i &= \xi^j\, \lambda'_h|^i{}_{,j}\\ &= \xi^j \sum_k (\cos\beta_{hk}\, \lambda_k|^i{}_{,j} + \lambda_k|^i\, \partial\cos\beta_{hk}/\partial x^j)\\ &= \sum_k (\cos\beta_{hk}\, \mu_k|^i + \lambda_k|^i\, \partial \cos\beta_{hk}/\partial s),\end{aligned} \tag{25}$$

where s is the arc of the curves C. And

$$\nu'_h|^i = \lambda'_h|^j\, \xi^i{}_{,j} = \sum_k \cos\beta_{hk}\, \nu_k|^i.$$

For $b'_h|^i$, the distantial spread vector of the curves C'_h and C, we have

$$b'_h|^i = \mu'_h|^i - \nu'_h|^i = \sum_k (\cos\beta_{hk}\, b_k|^i + \lambda_k|^i\, \partial\cos\beta_{hk}/\partial s). \tag{26}$$

The relations between the lengths of the vectors in question are given by

$$\begin{aligned}\sum_h 1/r'^2_h = \sum_h 1/r_h^2 + 2 \sum_{h,k,l} g_{ij}\, \mu_k|^i\, \lambda_l|^j \cos\beta_{hk}\, \partial\cos\beta_{hl}/\partial s\\ + \sum_{h,k} (\partial\cos\beta_{hk}/\partial s)^2,\end{aligned} \tag{27}$$

and

$$\begin{aligned}\sum_h 1/b'^2_h = \sum_h 1/b_h^2 + 2 \sum_{h,k,l} g_{ij}\, b_k|^i\, \lambda_l|^j \cos\beta_{hk}\, \partial\cos\beta_{hl}/\partial s\\ + \sum_{h,k} (\partial\cos\beta_{hk}/\partial s)^2.\end{aligned} \tag{28}$$

---

[11] Graustein, *loc. cit.*, p. 563.

The equations (25) through (28) can, of course, be regarded as the relations between the angular and distantial spread vectors of the n congruences of any two orthogonal ennuples with respect to an arbitrary congruence of curves C, without bringing in any notion of sheaves.

From (27) and (28) we conclude

THEOREM 17. *The sums of the squares of the lengths of the angular or distantial spread vectors of any orthogonal ennuples of the two sheaves are the same with respect to the curves along which the angles β_{hk} are constant.*[12]

In particular, if all the curves C_h of the ennuple of p are parallel with respect to a congruence of curves C along which the angles β_{hk} are constant, then all curves of both sheaves are parallel with respect to the curves C. A corresponding result can be stated for equidistance.

If the n congruences C_h are normal and consist of geodesics, then (25) becomes

$$\mu'_h|^i = \sum_k \lambda_k|^i \, \partial \cos \beta_{hk}/\partial s,$$

and

$$1/r'^2_{hk} = \sum_k (\partial \cos \beta_{hk}/\partial s)^2. \tag{29}$$

Hence,

THEOREM 18. *The square of the length of the angular spread vector of any congruence of curves C' in a Euclidean V_n with respect to the curves of any other congruence of curves C is equal to the sum of the squares of the directional derivatives in the direction of the curves C of the angles β_k between the curves C' and the curves C_k, $(k = 1, 2, \cdots, n)$, of an orthogonal ennuple of normal congruences of geodesics.*[13]

If the curves C belong to the sheaf p, and again the n congruences of curves C_h are normal and consist of geodesics, then the curves C are also geodesics, and (26) becomes

$$b'_h|^i = \mu'_h|^i = \sum_k \lambda_k|^i \, \partial \cos \beta_{hk}/\partial s,$$

and

$$1/b'^2_h = 1/r'^2_h = \sum_k (\partial \cos \beta_{hk}/\partial s)^2.$$

LAKE ERIE COLLEGE.

---

[12] A generalization of a result for angular spreads in a V_2 given by Graustein, "Parallelism and equidistance in classical differential geometry," *Transactions of the American Mathematical Society*, vol. 34 (1932), p. 570.

[13] This theorem is a generalization of Theorem 14 of Graustein, *loc. cit.*, p. 570.

ON THE NON-ALTERNATING IMAGES OF LINEAR GRAPHS.*

By Dick Wick Hall.

Let A and B be compact metric spaces and $T(A) = B$ a single valued continuous transformation. Then T is said to be *non-alternating*[1] provided that for no two distinct points x and y of B does the set $T^{-1}(x)$ separate the set $T^{-1}(y)$ in A, i. e., there exists no separation $A - T^{-1}(x) = A_1 + A_2$ where $A_1 \cdot T^{-1}(y) \neq 0 \neq A_2 \cdot T^{-1}(y)$. If for each $x \in B$ the set $T^{-1}(x)$ is connected, then T is said to be *monotone*.[2] A connected set M is said to be *cyclic* if it contains no cut point, i. e., if $M - x$ is connected for every x in M. If M be a locally connected continuum, and we shall always assume that it is, then a subset E of M will be called a *maximal cyclic subset* if and only if it is not a proper subset of any other cyclic subset of M. A subset E of M will be called a *cyclic element*[3] of M provided E is either (a) a maximal cyclic subset of M, (b) a cut point of M, (c) an end point[4] of M. A cyclic element containing more than one point is called a *true cyclic element*. An arc A is said to *span* a point-set M if A has its end points but no other points in common with M. Throughout this paper we shall assume that all the linear graphs mentioned are connected, and that all the point-sets considered are imbedded in a three dimensional Euclidean space, since we deal only with 1-dimensional sets and any such set is topologically contained in an E_3.

In this paper a study is made of the possible images of a linear graph under a non-alternating transformation. It is shown that: I: A necessary and sufficient condition that a cyclic curve C be the non-alternating image of

* Received July 23, 1936; Revised February 11, 1937.

[1] See G. T. Whyburn, *American Journal of Mathematics*, vol. 46 (1934), pp. 294-302. This paper will be referred to as W.

[2] This terminology has been suggested by C. B. Morrey. See his paper "The topology of path surfaces," *American Journal of Mathematics*, vol. 57 (1935), pp. 17-50.

[3] See Kuratowski and Whyburn, *Fundamenta Mathematicae*, vol. 16 (1930), pp. 305-350.

[4] A point p is an end point of a locally connected continuum M provided that M contains no simple arc having p as an interior point. See R. L. Wilder, *Fundamenta Mathematicae*, vol. 7 (1925), p. 358. For this particular definition of end point see G. T. Whyburn, *Transactions of the American Mathematical Society*, vol. 29 (1927), Theorem 12, p. 385.

a linear graph is that C be the sum of a finite number of simple arcs; II: If B be the non-alternating image of a linear graph A, then every true cyclic element of B is the sum of a finite number of simple arcs; III: Every curve C which is the sum of a finite number of simple arcs is the non-alternating image of a linear graph.

THEOREM 1. *If A be a linear graph and $T(A) = B$ is non-alternating, then every true cyclic element of B is the sum of a finite number of simple arcs.*

Proof. By $(W, 3.4)$, if E_b be any true cyclic element of B there exists a non-alternating transformation $W(A) = E_b$ such that none of the sets $W^{-1}(x)$ separate A; and by $(W, 3.5)$, there exists a true cyclic element E_a of A such that $W(E_a) = E_b$. Then none of the sets $W^{-1}(x)$ separate E_a, since any set which did so would also separate A which is impossible.

Now since E_a is a true cyclic element of a linear graph we may write $E_a = \sum_1^k \bar{A}_i$, where k is finite and $\bar{A}_i = a_i b_i$ is the closure of a free arc A_i. Then W is monotone on A_i, for all i. Otherwise, for some x in E_b, $W^{-1}(x)$ would separate E_a, which is impossible. Two cases may arise: (a) If $W(a_i) = W(b_i) = c_i$ for any i, then $\bar{A}_i$ maps into a single closed curve having only the point c_i in common with the rest of E_b. Consequently, since E_b is cyclic we must have $W(\bar{A}_i) = E_b$, a simple closed curve. Hence E_b is the sum of two simple arcs and the theorem follows. (b) If $W(a_i) \neq W(b_i)$ for any i, then W is monotone on $\bar{A}_i$ for all i. Thus $W(\bar{A}_i) = B_i$ is a simple arc, and $E_b = \sum_1^k B_i$, which is the theorem.

LEMMA 1. *If K be a cyclic curve such that there exists a linear graph H and a non-alternating transformation $T(H) = K$, and if A be any simple arc such that $K + A$ is cyclic, then there exists a simple arc B spanning H, and a non-alternating transformation $Z(H + B) = K + A$.*

Proof. We may assume that H is cyclic, by $(W, 3.5)$. Let a', a'' be the end points of A, and b', b'' any points of $T^{-1}(a')$ and $T^{-1}(a'')$ respectively. Let B be a simple arc spanning H and having b' and b'' as end points. Define Z so that it is identical with T on H, while on B it is a homeomorphism sending B into A and such that $Z(b') = a'$, and $Z(b'') = a''$. Then $Z(H + B) = K + A$.

Moreover, T is non-alternating. Otherwise, we could find two points x, y in $K + A$ such that $Z^{-1}(x)$ separated $Z^{-1}(y)$ in $H + B$. Now (1): x cannot lie in $A - K \cdot A$. Otherwise, $Z^{-1}(x)$, which is a single point since Z is one-to-one on $B - B \cdot H$, would separate the cyclic set $H + B$, which is impossible. (2) y cannot lie in $A - K \cdot A$, since Z is one-to-one on $H - B \cdot H$, and no single point can be separated. (3) Consequently, both x and y must lie in K. Since $Z^{-1}(x)$ separates $Z^{-1}(y)$ in $H + B$, it follows that $Z^{-1}(y)$ contains two points y' and y'' such that $Z^{-1}(x)$ separates y' and y'' in $H + B$. Now not both y' and y'' may lie in H, since T is non-alternating on this set. Hence we may assume that y', say, lies in $B - B \cdot H$. By the definition of $Z, B \cdot Z^{-1}(x)$ is a single point, hence $Z^{-1}(x)$ cannot separate y' from both b' and b'' say not from b'. Then, since $Z^{-1}(x)$ separates y' from y'', it must separate b' from y''. But we may assume that $y'' \epsilon H$, and hence $Z^{-1}(x)$ must separate two points of H. Thus $T^{-1}(x)$ must separate two points of H; and consequently, by $(W, 1.41)$, x is a cut point of $K + A$, which is a contradiction.

Therefore, T is non-alternating, and the lemma is proved.

THEOREM 2. *A necessary and sufficient condition that a cyclic curve C be the non-alternating image of a linear graph is that C be the sum of a finite number of simple arcs.*

Proof. Necessity: This is immediate from Theorem I. Sufficiency: Let $C = \sum_1^n A_i$, where n is finite and each A_i is a simple arc. Let the $2n$ end points of these simple arcs be denoted by $a_1, a_2, \cdots, a_{2n}$, where an end point is counted once for each arc of which it is an end point. Since C is cyclic, it contains a simple closed curve K_1 passing through a_1 and a_2. If K_1 does not contain all the points a_i, let a_j be any one of these points which it does not contain. Then, by the three point theorem [5] we may find a simple arc spanning K_1 and containing a_j. Thus we have found a cyclic linear graph containing a_1, a_2, a_j. Repeating this process a finite number of times we shall obtain a cyclic linear graph K which is a subset of C and which contains all of the points a_i. Thus $K + A_i$ is cyclic for every i. We now write $C = K + \sum_1^n A_i$, and the theorem follows at once by n applications of Lemma 1, and the addition of one of the arcs A_i to K at each step.

[5] See W. L. Ayres, *Bulletin Académie Polonaise Science et Lettres* (1928), pp. 127-142.

By a θ_n-*curve* we shall mean a curve expressible as the sum of $(n + 2)$ simple arcs having the same end points but otherwise disjoint by pairs.

Using the same construction as that employed in Theorem 2, the following lemma is immediate.

LEMMA 2. *Every cyclic curve C expressible as the sum of n simple arcs having the same end points is the non-alternating image of a θ_n-curve.*

It can be shown that no θ_2 curve is the image of a θ_1-curve under a non-alternating transformation. Consequently, since a θ_2-curve A is easily expressible as the sum of two simple arcs having as end points two points interior to different free arcs of A, it follows that the n in Lemma 2 cannot be reduced. For by $(W, 4.6)$ A is not the non-alternating image of a θ_0-curve, that is, of a simple closed curve, and hence not the non-alternating image of any θ_k-curve for $k < 2$.

Our next lemma will remove the restriction that C be cyclic.

LEMMA 3. *Every curve C which is the sum of n simple arcs A_i $(i = 1, 2, \cdots, n)$ having the same end points a and b is the non-alternating image of a θ_n curve.*

Proof. For notational reasons we give the proof for the case $n = 2$, since the general case follows in precisely the same way.

Since every arc in C joining a and b must be in every A-set [6] containing a and b it follows at once that $C = C(a, b)$, that is, C is a simple cyclic chain joining a and b.

Let K be the set of all points separating a and b in C. Then, since C is a simple cyclic chain, we may write $C = (K + a + b) + \Sigma C_i$, where each C_i is a true cyclic element of C.

Define a θ_2-curve $H = \sum_1^4 a'x_ib'$, where the $a'x_ib'$ are simple arcs having the same end points but otherwise disjoint by pairs. We shall prove that C is the image of H under a non-alternating transformation.

Let axb be any simple arc joining a and b in C. Then, by the definition of K, it follows that K is a subset of axb. Let $Z_i(a'x_ib') = axb$ be a homeomorphism defined on $a'x_ib'$ (for each i) and sending a' and b' into a and b, respectively. Let $K_i = Z_i^{-1}(K + a + b)$. Define a transformation Z of H

[6] For definitions of the new terms used see G. T. Whyburn, *American Journal of Mathematics*, vol. 50 (1928), pp. 167-194. In connection with A-sets see also Kuratowski and Whyburn, *loc. cit.*

into a new curve H' as follows: (i) Z is identical with Z_i on K_i, (ii) Z is a homeomorphism on $H - \sum_1^4 K_i$. Then H' will coincide with C at all points of $(K + a + b)$. Moreover, by the definition of Z every true cyclic element of H' will consist of four simple arcs disjoint by pairs except for their end points and joining two points of $(K + a + b)$ in C. Thus every true cyclic element of H' will be a θ_2-curve. But every true cyclic element of C is the sum of two simple arcs having their end points in common and hence, by Lemma 2, every such true cyclic element is the non-alternating image of a θ_2-curve. From the proof of Theorem 2 it follows that we may pick non-alternating transformations sending the true cyclic elements of H' into those of C in such a manner that they will map the points of $H'_i \cdot K$ (where H'_i is the given true cyclic element) into themselves. Thus we may define a transformation $W(H') = C$ to be the identity transformation on K, and to send each true cyclic element of H' into the corresponding true cyclic element C_i of C (namely the one containing the same points of $(K + a + b)$) in a non-alternating manner.

Let H'_i be any true cyclic element of H' and suppose that $W(H'_i) = C_i$. Then, by (W, 1.41), since W is non-alternating on H'_i and C_i is cyclic, it follows that for no x in C_i does $W^{-1}(x)$ separate H'_i. Also, by definition of W, the images of the end points of each free arc of H'_i are distinct. Thus W is monotone on the closure of each free arc of H'_i.

Let axb be any simple arc joining a and b in H'. Then W is monotone on axb. To prove this it is sufficient to show that if p and q be any two points on axb such that $W(p) = W(q)$, and if z be any point between p and q on axb, then $W(z) = W(p) = W(q)$. Since W is the identity transformation on K, p and q cannot both lie in K, so we may assume that p lies on a free arc of some true cyclic element H'_j of H'. Then if q lies in H'_j our assertion is established since W is monotone on the closure of each free arc of this set. If q is not in H'_j, then, since W maps disjoint cyclic elements of H' into disjoint cyclic elements of C, it follows that q lies in some H'_k, where $H'_j \cdot H'_k = y$, a single point (which may or may not be q). Let pyq be a simple arc joining p to q in $H'_j + H'_k$. Then, since W is monotone on the closure of each free arc of both H'_j and H'_k, we have $W(pyq) = W(y)$, so that in particular, $W(z) = W(y) = W(p) = W(q)$, which proves W monotone on axb.

By definition the transformation $Z(H) = H'$ is monotone on the closure of each free arc of H, hence by (W, 2.2), if we define a transformation $T = WZ$ it follows that T is monotone on the closure of each free arc of H.

By definition we have $T(H) = W(H') = C$, so that the lemma will be established if we show that T is non-alternating on H.

The set H is locally connected, hence by $(W, 1.5)$, in order to show that $T(H) = C$ is non-alternating, it is sufficient to show that for each point $q \in C$ and each component K of $C - q$, the set $T^{-1}(K)$ is connected. Letting q be any point of C, two cases must be considered.

Case 1. q is a cut point of C.

Then q is distinct from both a and b. Recalling that by definition $H = \sum_1^4 a'xb'$, it follows that $B_i \equiv T(a'x_ib') = ax_ib$ is a simple arc since T is monotone on the closure of each of the simple arcs $a'x_ib'$. Since q cuts C it lies on all the arcs B_i, so we may write $B_i = ax_iq + qy_ib$. Then since $C = \sum_1^4 B_i$, we have $C - q = \sum_1^4 (ax_iq - q) + \sum_1^4 (qy_ib - q) \equiv M_1 + M_2$. Then M_1 and M_2 are both connected and closed in $C - q$. Consequently, since $C - q$ is disconnected, these sets are mutually separated, and hence components of $C - q$.

Let a transformation T_i be defined as identical with T on $a'x_ib'$, and undefined elsewhere. Then T_i is monotone and $T(a'x_ib' = B_i$. Thus for every i the sets $T_i^{-1}(ax_iq - q)$ and $T_i^{-1}(qy_ib - q)$ are connected. But these are precisely the sets

$$T^{-1}(ax_iq - q) \cdot (a'x_ib'), \tag{1}$$

$$T^{-1}(qy_ib - q) \cdot (a'x_ib'), \tag{2}$$

respectively, so that the sets (1) and (2) are connected for all i. Thus

$$a'x_ib' - T^{-1}(q) = T^{-1}(ax_iq - q) \cdot (a'x_ib') + T^{-1}(qy_ib - q) \cdot (a'x_ib'),$$

and therefore,

$$\sum_1^4 a'x_ib' - T^{-1}(q) = \sum_1^4 T^{-1}(ax_iq - q) \cdot (a'x_ib') + \sum_1^4 T^{-1}(qy_ib - q) \cdot (a'x_ib'),$$

so that

$$H - T^{-1}(q) = \sum_1^4 T^{-1}(ax_iq - q) \cdot (a'x_ib') + \sum_1^4 T^{-1}(qy_ib - q) \cdot (a'x_ib') \equiv N_1 + N_2,$$

and from (1) and (2), it follows that N_1 and N_2 are connected. They are evidently disjoint and closed in their sum, so they are mutually separated.

For any point $x \in N_1$, we have $T(x) \in \sum_1^4 (ax_iq - q) \subset M_1$, so that $x \in T^{-1}(M_1)$, hence $N_1 \subset T^{-1}(M_1)$, and similarly $N_2 \subset T^{-1}(M_2)$.

Moreover, for any point $x \in T^{-1}(M_1)$ we have $T(x) \in M_1$ so that $T(x)$ non $\in M_2$. Consequently, since $N_2 \subset T^{-1}(M_2)$, we have x non $\in N_2$; therefore $x \in N_1$, so that $T^{-1}(M_1) \subset N_1$. Thus $N_1 = T^{-1}(M_1)$, and similarly $N_2 = T^{-1}(M_2)$, so both of these sets are connected and the lemma follows for Case 1.

Case 2. q is a non-cut point of C.

We have again $B_i = T(a'x_ib')$. If $q \in B_i$ write $B_i = ax_ib = ax_iq + qy_ib$. Otherwise, write $B_i = ax_ib$. Let the $a'x_ib'$ be numbered so that for some integer j, $q \in B_i$, $(1 \leq i \leq j)$, q non $\in B_i$, $(i > j)$. (If $q \in B_i$ for every i, then we have, of course, $j = 4$ since H contains but four free arcs by hypothesis). Then

$$C - q = \sum_{j+1}^{4} ax_ib + \sum_{1}^{j} (ax_iq - q) + \sum_{1}^{j} (qy_ib - q).$$

Define

$$M = \sum_{1}^{j} T^{-1}(ax_iq - q) \cdot (a'x_ib'), \qquad N = \sum_{1}^{j} T^{-1}(qy_ib - q) \cdot (a'x_ib'),$$
$$Z = \sum_{j+1}^{4} T^{-1}(ax_ib) \cdot (a'x_ib').$$

By the reasoning used in Case 1, each of the sets M, N, and Z is vacuous or connected.

Assume, first, that Z is non-vacuous. Then q is distinct from both a and b so that M contains a', N contains b', and Z contains both of these points. Therefore, since each of these sets is connected we have that $(M + N + Z)$ is a connected set.

Secondly, if Z is vacuous, then q occurs on every simple arc B_i. If $q = a$, then $M = 0$, and both N and Z contain b', so that $(M + N + Z)$ is connected. Similarly, if $q = b$, then $N = 0$, both M and Z contain a', and, consequently, $(M + N + Z)$ is connected. If q is distinct from both a and b, then, since q does not cut C, there exist two simple arcs B_j and B_k in C and a point p in C which precedes q on the arc B_k but follows it on the arc B_j. Then M and N have $T^{-1}(p)$ in common so that $(M + N)$ and hence $(M + N + Z)$ is connected. Therefore, in every event, $(M + N + Z)$ is a connected set.

But, since $C = \Sigma\, T(a'x_ib')$, we have that $T^{-1}(C - q) = M + N + Z$, so that this set is connected. Therefore, for any point q of C and every component K of $C - q$, the set $T^{-1}(K)$ is connected. Consequently, by $(W, 1.5)$, the transformation $T(H) = C$ is non-alternating, as was to be proved.

LEMMA 4. *Suppose that A, C, and $A + C$ are connected and that $A \cdot C = p$, a single point. Let $T'(A) = B$ and $T^2(C) = D$ be non-alternating*

transformations such that $T' = T^2$ on AC. Define a transformation $T = T'$ on A, $T = T^2$ on C. Then a necessary and sufficient condition that T be non-alternating on $(A + C)$ is that $B \cdot D = T(p)$.

Proof. The condition is clearly necessary. Otherwise, the inverse of some point x of $(B + D) - p$ would intersect both $A - p$ and $C - p$, so that this inverse would be separated by $T^{-1}T(p)$ in $(A + C)$ which would make T alternate on this set.

The condition is also sufficient. Otherwise, there exist two points y', y'' in $(A + C)$, with $T(y') = T(y'')$, and a point x in $B + D$ such that $T^{-1}(x)$ separates y' from y'' in $(A + C)$. Then (i) if y' and y'' both lie in A or C, we have a contradiction to the fact that T is non-alternating on these respective sets; (ii) if $y' \in A$, $y'' \in C$, or conversely, then, since $T(y') = T(y'')$ and $B \cdot D = p$, we have $T(y') = T(y'') = T(p)$, and thus p non $\in T^{-1}(x)$. Hence, by the definition of T, $T^{-1}(x)$ cannot separate either y' or y'' from p; so that it cannot separate y' from y'', and the lemma follows.

LEMMA 5. *If $C = \sum_1^n A_i$ be connected, where n is finite and each A_i is a connected A-set (in some locally connected continuum S) which is the non-alternating image of a linear graph and where for no i, j $(i \neq j)$ does $A_i \cdot A_j$ contain more than one point, then C is the non-alternating image of a linear graph.*

Proof. By hypothesis there exists a set of disjoint linear graphs H_i $(i = 1, 2, \cdots, n)$ and a set of non-alternating transformations $T_i(H_i) = A_i$.

Consider the A-set A_1. Since C is connected, at least one of the other sets A_i, say A_2, must intersect it. Then $A_1 \cdot A_2 = p_{12}$, a single point by hypothesis. Let p_1, q_1 be any two points of $T_1^{-1}(p_{12})$ and $T_2^{-1}(p_{12})$, respectively, and translate H_2 until it has the single point $p_1 = q_1$ in common with H_1. Then, by Lemma 5, the transformation T_{12} which is T_1 on H_1 and T_2 on H_2 is non-alternating, and $T_{12}(H_1 + H_2) = A_1 + A_2$.

Repeating this process we may add on one A-set at a time, and at each stage secure a non-alternating transformation sending $\sum_1^k H_i$, for example, into $\sum_1^k A_i$ at the k-th stage. It follows at once that A_k cannot have two points in common with $\sum_1^{k-1} A_i$, since if p and q were any two such points they would, by hypothesis, lie in different A-sets of the above sum. Then, using the fact that each of the A-sets is connected and locally connected, we could construct a simple arc joining p and q and not lying wholly in A_k, which is contradictory

to the definition of an A-set. Thus the extension may be made in exactly the same way at every stage and the lemma follows.

LEMMA 6. *Every simple cyclic chain which is the sum of a finite number of simple arcs is the non-alternating image of a linear graph.*

Proof. Let $C(a, b)$ be the simple cyclic chain which is the sum of j simple arcs by hypothesis, and let a_i $(i = 1, 2, \cdots, 2j)$ be the endpoints of these simple arcs. Then we may find $2j$ cyclic elements of $C(a, b)$ whose sum contains all the points a_i. Let n be the number of these cyclic elements which are distinct, and number them K_i in the order in which they occur from a to b. Since $C(a, b)$ is a simple cyclic chain, let it be expressed in the form $C = (K + a + b) + \Sigma C_i$, where the C_i are its true cyclic elements, and K consists of all those points separating a and b in $C(a, b)$. Let axb be any simple arc joining a and b in $C(a, b)$, and let x_i, y_i be respectively its first and last intersections with the set K_i. Evidently $x_i = y_i$ if K_i is degenerate. Then $x_1 = a$, $y_n = b$, while all of the points x_i, y_i which are distinct from these two are points of K. Consequently, the choice of the points x_i, y_i is independent of the arc axb we used. Define $M_{2i-1} = C(x_i, y_i)$, $M_{2i} = C(y_i, x_{i+1})$ as simple cyclic chains in C. Some of the sets M_i may be degenerate.

Evidently the chains M_i as constructed are finite in number and $C(a, b) = \sum_1^{2n-1} M_i$. Moreover, these sets are A-sets, by definition, and for any i, j $(i \neq j)$, $M_i \cdot M_j$ contains not more than one point. Our lemma will thus follow by Lemma 6 if we show that, for every i, M_i is the non-alternating image of a linear graph. This follows at once for every M_{2i-1} by Theorem II, since every such set is a K_i, hence cyclic, and is the sum of a finite number of simple arcs, since C is. Now M_{2i} is a cyclic chain joining y_i to x_{i+1}, and both these points belong to K. Moreover, with the possible exception of these two points, M_{2i} cannot contain end points of any of the simple arcs which go to make up $C(a, b)$; whence M_{2i} is the sum of a finite number of simple arcs joining y_i and x_{i+1}. The lemma is then an immediate consequence of Lemma 3.

THEOREM III. *Every curve C which is the sum of a finite number of simple arcs is the non-alternating image of a linear graph.*

Proof. Let C be the sum of j simple arcs. Then these arcs have not more than $2j$ distinct end points; whence C contains not more than $2j$ nodes,[7] since every node of C must evidently contain an end point of at least one of the j

[7] See G. T. Whyburn, *American Journal of Mathematics*, vol. 50 (1928), p. 178.

simple arcs of which C is the sum. Let there be h nodes in C and call them E_i $(i = 1, 2, \cdots, h)$. Let p_i be any non-cut point contained in E_i. Define $M_2 = C(p_1, p_2)$. Let p_1xp_3 be any simple arc joining p_1 to p_3 and let q_3 be the last intersection of this arc with M_2. Define $M_3 = C(p_3, q_3)$. In general, if M_{k-1} has been defined let p_1xp_k be a simple arc joining p_1 to p_k in C and let q_k be its last intersection with $\sum_2^{k-1} M_i$. Define $M_k = C(p_k, q_k)$.

Evidently, $C = \sum_2^h M_i$, and each M_i is the non-alternating image of a linear graph by virtue of Lemma 6. Moreover, the conditions of Lemma 5 are obviously fulfilled so that the theorem follows by virtue of that lemma.

The converse of Theorem III is false. For, let D be any dendrite which is not the sum of a finite number of simple arcs (and such a dendrite may easily be constructed). Then, by (W, p. 301), D is a boundary curve, and hence, by (W, 4.6), D is the non-alternating image of a circle, which is certainly a linear graph. Therefore, the non-alternating image of a circle need not be the sum of a finite number of simple arcs.

UNIVERSITY OF VIRGINIA.

REPRESENTATIONS IN CERTAIN PURE FORMS OF DEGREES HIGHER THAN THE SECOND.*

By E. T. Bell.

1. Introduction. The method of Lagrange[1] for finding parametric solutions of certain diophantine equations may be taken as the point of departure for obtaining representations of integers in special forms of degree $\geqq 2$. This will be illustrated by starting from the equation

$$(1.1) \qquad a^3 + b^3 + pc^3 = 0, \qquad pabc \neq 0,$$

in which p is a constant integer. Numerical examples are given in § 5. Another method is indicated in § 6, with examples. *Throughout the paper, x, y, z, u, v, w denote real or complex variables, other small letters, rational integers.* Without loss of generality a, b may be taken coprime, $(a, b) = 1$, and any divisor r^3 of p may be absorbed in c if desired. It was shown by Lucas[2] that if $\{d\}$ denotes d divided by the greatest cube divisor of d, a necessary and sufficient condition that (1.1) have a solution is that p be of the form $\{st(s+t)\}$.

If $(a, b, c) = (a_n, b_n, c_n)$ is any solution of (1.1), $(a_{n+1}, b_{n+1}, c_{n+1})$ is also a solution, where

$$(1.2) \quad a_{n+1} = a_n(a_n^3 + 2b_n^3), \quad b_{n+1} = -b_n(2a_n^3 + b_n^3), \quad c_{n+1} = c_n(a_n^3 - b_n^3),$$

and if $(a_{n+1}, b_{n+1}, c_{n+1}) \neq (ka_n, kb_n, kc_n)$, the two solutions are said to be distinct.[3] A parametric solution of (1.1), due to Lucas, is

$$(1.3) \quad \begin{aligned} a &= x^3 - y^3 + 6x^2y + 3xy^2, & b &= -x^3 + y^3 + 3x^2y + 6xy^2, \\ c &= 3(x^2 + xy + y^2), & -p &= xy(x+y); \end{aligned}$$

if a, b, c, p are integers, x, y run through all integers. By means of (1.3),

* Received April 30, 1937.

[1] Supplement to Euler's *Algebra*; see also R. D. Carmichael's *Diophantine Analysis*.

[2] E. Lucas, *American Journal of Mathematics*, vol. 2 (1879), p. 184. The paper by L. Holzer in *Journal für Mathematik*, vol. 159 (1928), pp. 93-100, discusses the same equation, and obtains (incidentally) some of the results of Lucas and Sylvester, whose papers on the subject appear to have been overlooked by the author.

[3] Lucas asserts, *loc. cit.*, p. 185, that if $|a| \neq |b|$, there is an infinity of distinct solutions. This has not been proved.

Lucas solved (in particular) (1.1) with $p = 6$, which Legendre had mistakenly asserted to be unsolvable in integers.

2. Identities from (1.1). Let (a, b, c) be any solution of (1.1), without the conditions $(a, b) = 1$, $|c|$ free of cube divisors > 1, so that

$$(2.1) \qquad a^3 + b^3 + pc^3 = 0, \qquad pabc \neq 0.$$

We shall determine the parameters w, x, y, z so that, identically in w, x, y, z,

$$(2.2) \qquad (aw + x)^3 + (bw + y)^3 + p(cw + z)^3 \equiv 3Aw + B,$$

with A, B independent of w.

LEMMA 1. *If x, y, z are such that*

$$(2.3) \qquad a^2x + b^2y + pc^2z = 0,$$

where (a, b, c) is a particular solution of (2.1), then identically in w, x, y, z,

$$(2.4) \qquad pc^4[(aw + x)^3 + (bw + y)^3 + p(cw + z)^3] \equiv -(bx - ay)^2(3abcw + bcx + cay + abz).$$

For, from (2.1), (2.2) we get (2.3) and $A = ax^2 + by^2 + pcz^2$, $B = x^3 + y^3 + pz^3$; whence, eliminating z by (2.3) we have

$$-pc^3A = ab(bx - ay)^2,$$
$$p^2c^6B = (bx - ay)^2[b(2a^3 + b^3)x + a(a^3 + 2b^3)y].$$

Hence, if (2.2) be multipled throughout by p^2c^6, the right becomes $(bx - ay)^2F$,

$$\begin{aligned} F &\equiv -[3abpc^3w - (2a^3 + b^3)bx - (a^3 + 2b^3)ay]. \\ F &= -[3abpc^3w - (a^3 - pc^3)bx - (b^3 - pc^3)ay], \\ &= -[pc^3(3abw + bx + ay) - ab(a^2x + b^2y)], \\ &= -pc^2(3abcw + bcx + cay + abz); \end{aligned}$$

which completes the proof of (2.4).

LEMMA 2. *If, without loss of generality, $(a, b) = 1$ in (2.1), so that (integers) r, s may be determined such that*

$$(2.5) \qquad a^2r + b^2s = -1,$$

then, identically in w, u, v,

$$(2.6) \qquad (aw + prc^2u + b^2v)^3 + (bw + psc^2u - a^2v)^3 + p(cw + u)^3 \equiv -pM^2(3abcw + R),$$

where $M \equiv (br - as)u - cv$,

$$R \equiv (prbc^3 + psac^3 + ab)u + (b^3 - a^3)v.$$

For, from (2.5), (2.3) we get

$$(x, y, z) = (prc^2u + b^2v, psc^2u - a^2v, u),$$

with (u, v) arbitrary, and the result follows by (2.1), (2.4).

In (2.6) we now choose $(u, v) = (m, n)$, m n integers, and reduce the right numerically. If $[y]$ is the greatest integer in y, we write

$$(2.7) \qquad G \equiv \left[\frac{|R|}{3\,|abc|}\right], \qquad \eta \equiv \operatorname{sgn} R, \qquad \epsilon \equiv \operatorname{sgn}(abcR)$$

after having replaced (u, v) by (m, n) in R, where $\operatorname{sgn} y$ denotes $1, 0, -1$ according as $y > 0$, $y = 0$, $y < 0$. Then

$$(2.8) \qquad |R| = 3\,|abc|\,G + \rho, \qquad 0 \leqq \rho < 3\,|abc|.$$

Replacing w by $w - \epsilon G$ in $3\,abcw + R$ we find $3abcw + \eta\rho$.

Lemma 3. *With* G, ρ *as in* (2.7), (2.8),

$$(aw + prc^2m + b^2n - \epsilon aG)^3 + (bw + psc^2m - a^2n - \epsilon bG)^3$$
$$+ p(cw + m - \epsilon G)^3 \equiv -pM^2(3abcw + \eta\rho),$$

identically in w, *where*

$$M \equiv (br - as)m - cn,$$

and a, b, c, r, s *are as in Lemma* 2.

Returning to (2.6) we recall that (r, s) are defined by (2.4). Let

$$(2.9) \qquad (br - as, c) = g.$$

Then integers h, k may be found such that

$$(2.10) \qquad (br - as)h - ck = g,$$

and (2.12) is a solution of

$$(2.11) \qquad (br - as)u - cv = gx,$$

$$(2.12) \qquad u = hx + cy, \qquad v = kx + (br - as)y.$$

LEMMA 4. *With r, s, g, h, k as in* (2.4), (2.9), (2.10),

$$-pg^2x^2(3abcw + R) \equiv X^3 + Y^3 + pZ^3$$

identically in w, x, y, where

$$\begin{aligned} X &\equiv aw + (prc^2h + b^2k)x + ay, \\ Y &\equiv bw + (psc^2h - a^2k)x + by, \\ Z &\equiv cw + hx + cy, \\ R &\equiv [h(prbc^3 + psac^3 + ab) + k(b^3 - a^3)]x \\ &\qquad + [c(prbc^3 + psac^3 + ab) + (br - as)(b^3 - a^3)]y, \end{aligned}$$

and a, b, c are as in Lemma 2.

As in Lemma 2, R in Lemma 4 may be reduced numerically when x, y are integers. Write

$$\begin{aligned} \mu &\equiv h(prbc^3 + psac^3 + ab) + k(b^3 - a^3) \\ \sigma &\equiv c(prbc^3 + psac^3 + ab) + (br - as)(b^3 - a^3); \end{aligned} \tag{2.13}$$

μ, σ are integers, and in Lemma 4, $R = \mu x + \sigma y$.

3. Consequences of Lemmas 3, 4. In Lemma 3 take $w = 0$. Then

THEOREM 1. *With M as in Lemma 3, every $\pm p\rho M^2$ (integer) is of the form $\alpha^3 + \beta^3 + p\gamma^3$, with α, β, γ integers, and with at most 3 exceptions $M \neq 0$, all of α, β, γ may be chosen $\neq 0$.*

In Lemma 4 take $w = 0$; $y = 0$. Then

THEOREM 2. *Every $\pm pg^2\sigma x^2 y$, and every $\pm 3abcpg^2x^2w$ is of the form*

$$X^3 + Y^3 + p(Z^3 + g^2\mu U^3),$$

with σ, μ as in (2.13), *g as in* (2.9). *If x, y, w are integers, X, Y, Z, U may be chosen integers, and in each case with at most 4 exceptions, all different from zero.*

4. Further consequences of (2.3). Returning to Lemma 1, we now find common solutions (x, y, z) of (2.3) and

$$bx - ay = hpc. \tag{4.1}$$

Assuming (without loss of generality) as before that $(a, b) = 1$, we can find f, g such that

$$(4.2) \qquad bf - ag = 1.$$

Then the solution of (4.1) is

$$(x, y) = (hpcf + ka, hpcg + kb),$$

and this will give z an integer in (2.3) provided

$$pc^2 \mid (hpcf + ka), \qquad pc^2 \mid (hpcg + kb).$$

Hence

$$p \mid k, \quad k = tp; \qquad c \mid h, \quad c^2 \mid t; \qquad h = mc, \quad t = nc^2,$$

and a common solution of (2.3), (4.1) is

$$(4.3) \qquad \begin{aligned} x &= pc^2(an + fv), \qquad y = pc^2(bn + gv), \\ z &= pc^3n - (a^2f + b^2g)v. \end{aligned}$$

Replacing w by $w - pc^2n$ in (2.4) and reducing the result we find

LEMMA 5. *If (a, b, c) is any solution of (2.1) with $(a, b) = 1$, and f, g are determined by (4.2), then, identically in w, v,*

$$(4.4) \qquad \begin{aligned} (aw + pc^2fv)^3 + (bw + pc^2gv)^3 + p[cw - (a^2f + b^2g)v]^3 \\ \equiv -pv^2[3abcw - \{a^3 - (3ag + 1)pc^3\}v]. \end{aligned}$$

For numerical reductions of (4.4) we write

$$(4.5) \qquad \begin{gathered} A \equiv 3abc, \qquad B \equiv (3ag + 1)pc^3 - a^3, \\ |B| = Q\,|A| + R, \qquad 0 \leqq Q, \qquad 0 \leqq R < |A|; \\ \operatorname{sgn}(AB) \equiv \epsilon, \qquad \operatorname{sgn} B \equiv \eta. \end{gathered}$$

LEMMA 6. *With the notations of Lemma 5 and (4.5),*

$$(aw + \alpha v)^3 + (bw + \beta v)^3 + p(cw + \gamma v)^3 \equiv -pv^2(Aw + \eta Rv),$$

identically in w, v, where

$$\alpha \equiv pc^2f - \epsilon aQ, \qquad \beta \equiv pc^2g - \epsilon bQ, \qquad -\gamma \equiv a^2f + b^2g + \epsilon cQ.$$

Since every integer n is of the form rs^2 in at least one way we have

THEOREM 3. *Every $3abcpn$ is of the form*

$$\alpha^3 + \beta^3 + p(\gamma^3 + \eta R\delta^3)$$

with the notation as in Lemma 6; and with at most 3 exceptions n, all the integers $\alpha, \beta, \gamma, \delta$ may be chosen $\neq 0$.

THEOREM 4. *In the statement of Theorerm 2, x^2y and x^2w may be replaced by n.*

5. Identities with fourth powers. By integrating the identities in the preceding lemmas with respect to the parameters between suitable limits we ascend from identities involving cubes to others involving fourth powers. It will be sufficient to illustrate the general process for Lemma 6; the actual ascent in numerical examples is most readily made directly from the examples. Integrating with respect to w between the limits 0 and w we find

LEMMA 7. *Identically in w, v,*

$$bc(aw+\alpha v)^4+ca(bw+\beta v)^4+pab(cw+\gamma v)^4 \\ -(bc\alpha^4+ca\beta^4+pab\gamma^4)v^4 \equiv -2pv^2w(Aw+2\eta Rv),$$

the notation being as in Lemma 6.

Integration with respect to v between the limits 0 and v gives

LEMMA 8. *Identically in w, v,*

$$\begin{aligned}3[\beta\gamma(aw+\alpha v)^4+\gamma\alpha(bw+\beta v)^4+p\alpha\beta(cw+\gamma v)^4 \\ -(\beta\gamma a^4+\gamma\alpha b^4+p\alpha\beta c^4)w^4] \\ \equiv -pv^3(4Aw+3\eta Rv),\end{aligned}$$

the notation being as in Lemma 6.

For $w=1$ or $v=1$ the last two give theorems for fourth powers similar to those for cubes.

5. Numerical examples. An indefinite number of special results are furnished by the preceding lemmas and theorems for particular solutions of (2.1). It will suffice to illustrate Lemma 6 (a further, more systematic solution from the numerous results on hand will be given on another occasion). The obvious solution $(a,b,c,p)=(a,b,-1,a^3+b^3)$ gives some interesting results for various a, b.

The choice $(a,b,c,p)=(1,1,-1,2)$, $(f,g)=(1,0)$, gives $A=-3$, $B=-3$, $Q=1$, $R=0$, $\epsilon=1$, $\eta=-1$, $\alpha=1$, $\beta=-1$, $\gamma=0$. Hence (by Lemma 6),

$$(5.1)\qquad (w+v)^3+(w-v)^3-2w^3=6wv^2,$$

a well known identity.

(5.2) *Every $6n$ is of the form $a^3+b^3-2c^3$, and if $n>1$, we may choose $a, b, c>0$.*

The iteration (1.2) applied to $(a, b, c)=(1, 1, -1)$ gives

$$(a, b, c)=(3, -3, 0),$$

not a solution of (2.1) since here $pabc=0$. For

$$(a, b, c, p)=(2, -1, -1, 7)$$

we take $(f, g)=(-3, 1)$ and find $A=6$, $B=-57$, $Q=9$, $R=3$, $\epsilon=-1$, $\eta=1$, $\alpha=-3$, $\beta=-2$, $\gamma=2$:

(5.3) $(2w-3v)^3-(w+2v)^3-7(w-2v)^3=-21v^2(2w-v).$

Replace w by $-w$. Then

(5.4) *Every $42n$ is of the form $a^3-b^3+7(c^3-3d^3)$, and if $n>2$, we may take $a, b, c, d>0$.*

Iteration as in (1.2) of $(a, b, c)=(2, -1, -1)$ gives the new solution $(a, b, c)=(4, 5, -3)$ of (1.2) with $p=7$. For this we find $(f, g)=(1, 1)$, $A=-180$, $B=-2521$, $Q=14$, $R=1$, $\epsilon=1$, $\eta=-1$, $\alpha=7$, $\beta=-7$, $\gamma=1$:

(5.5) $(4w+7v)^3+(5w-7v)^3-7(3w-v)^3\equiv 7v^2(180w+v).$

(5.6) *Every $1260n$ is of the form $a^3+b^3-7(c^3+d^3)$, and if $n>1$, we may take $a, b, c, d>0$; by (5.4) every $1260n$ is also of the form*

$$a^3-b^3+7(c^3-3d^3)$$

with $a, b, c, d>0$.

In the same way we find the following. From $(a, b, c, p)=(2, 1, -1, 9)$,

(5.7) $(2w+3v)^3+(w-3v)^3-9(w+v)^3=9v^2(6w-v)$;

from $(a, b, c, p)=(2, 3, -1, 35)$,

(5.8) $(2w+7v)^3+(3w-7v)^3-35(w-v)^3=35v^2(18w+v)$;

from $(a, b, c, p)=(3, -2, -1, 19)$,

(5.9) $(2w+5v)^3-(3w-2v)^3+19(w-2v)^3=19v^2(18w-v)$;

from $(a,b,c,p)=(5,-4,-1,61)$,

(5.10) $(4w+13v)^3-(5w+v)^3+61(w-3v)^3=183v^2(20w+3v)$.

From (5.7)–(5.10) we write down the results corresponding to (5.6), etc. Thus from (5.10),

(5.11) *Every* $3660n$ *is of the form* $a^3-b^3+61(c^3-9d^3)$, *and if* $n>3$, *we may take* $a,b,c,d>0$; *every* $183(20n+3)$ *is of the form* $a^3-b^3+61c^3$, *and if* $n>3$, $a,b,c>0$.

One example of § 4 will suffice. Integrating (5.7) with respect to v between 0 and v we get

(5.12) $(2w+3v)^4-(w-3v)^4+3[4w^4+9v^4-9(w+v)^4]=216wv^3$;

(5.13) *Every* $27(8n-1)$ *is of the form* $a^4-b^4+3(4c^4-9d^4)$, *with* $a,b,c,d>0$ *if* $n>3$.

We give some miscellaneous examples, illustrative of general devices. Taking $v=\pm(1,2,3)$ in

$$(20w-3v)^3-(17w-3v)^3-9(7w-v)^3=9v(6w-v)^2,$$

obtained by the preceding methods, we get

$$\begin{aligned}(5.14)\quad &(20w-3)^3-(17w-3)^3-9(7w-1)^3=9(6w-1)^2,\\ &8(10w-3)^3-(17w-6)^3-9(7w-2)^3=72(3w-1)^2,\\ &(20w-9)^3-(17u-9)^3-9(7u-3)^3=243(2w-1)^2,\\ &64(5w-3)^3-(17w-12)^3-9(7w-4)^3=144(3w-2)^2,\\ &125(4w-3)^3-(17w-15)^3-9(7w-5)^3=45(6w-5)^2,\\ &8(10w-9)^3-(17w-18)^3-9(7w-6)^3=1944(w-1)^2.\end{aligned}$$

Hence, for example, *every* $72(3n-1)^2$ *is of the form* $8a^3-b^3-9c^3$, *with* $a,b,c>0$ *if* $n>0$. An interesting specimen of this kind, from another identity, is

(5.15) *Every* $168n^2$ *is of the form* $a^3+8b^3-7c^3$, *with* $a,b,c>0$ *if* $n>0$, *and every* $21(2n+1)^2$ *is of the form* $a^3+b^3-7c^3$, *with* $a,b,c>0$ *if* $n\geqq 0$.

Another kind is illustrated from the pair

$$(2w+3v)^3+(w-3v)^3-9(w+v)^3=9v^2(6w-v),$$
$$(2w-7v)^3+(3w+7v)^3-35(w+v)^3=35v^2(18w-v).$$

In the first replace w by $3w$ and subtract from the second. Then

$$35[3(2w+v)^3+3(w-v)^3-(3w+v)^3+(w+v)^3]$$
$$=(2w-7v)^3+(3w+7v)^3.$$

In this we now make any term, say $(3w+7v)^3$, equal to x^3. Hence (in this case) $w=-2x-7u$, $v=x+3u$, and we get

(5.16) $$x^3=(11x+35u)^3+35[(5x+18u)^3 - (x+4u)^3-3(3x+10u)^3-3(3x+11u)^3];$$

(5.17) *Every* n^3 *is of the form* $a^3+35(b^3-c^3-3d^3-3e^3)$, *and if* $n\neq 0$, *all of* $a,\cdots,e$ *may be chosen* >0 *in an infinity of ways.*

Integration of (5.16) gives

(5.18) $$11x^4=(11x+35u)^4+7[11(5x+18u)^4+224u^4] - 385[(x+4u)^4+(3x+10u)^4+(3x+11u)^4];$$

and hence, on replacing u by $11u$,

(5.19) $$x^4=1331(x+35u)^4+7[(5x+198u)^4+42592u^4] - 35[(x+44u)^4+(3x+110u)^4+(3x+121u)^4];$$

(5.20) *Every* n^4 *is of the form*

$$1331a^4+7(b^4+42592c^4)-35(d^4+e^4+f^4),$$

and all of $a,\cdots,f$ *may be chosen* >0 *in an infinity of ways if* $n\neq 0$.

Differentiation of (5.16) with respect to x gives

(5.21) $$x^2=11(11x+35u)^2+35[5(5x+18u)^2 - (x+4u)^2-9(3x+10u)^2-9(3x+11u)^2];$$

(5.22) *Every* n^2 *is of the form* $11a^2+35(5b^2-c^2-9d^2-9e^2)$, *and if* $n\neq 0$, *all of* $a,\cdots,e$ *may be chosen* >0 *in an infinity of ways.*

Differentiating (5.16) with respect to u, and replacing x by $16w-35v$, u by $-5w+11v$ in the result, gives

$$(5.23)\quad w^2 = 4(4w - 9v)^2 + 3[10(2w - 5v)^2 + 11(7w - 16u)^2 - 6(10w - 23v)^2];$$

(5.24) *Every* n^2 *is of the form* $4a^2 + 3(10b^2 + 11c^2 - 6d^2)$, *and if* $n \neq 0$, *all of* $a, \cdots, d$ may be chosen > 0 *in an infinity of ways.*

Differentiation of the last of (5.14) gives

$$(5.25)\quad 80(10w + 1)^2 - 17(17w - 1)^2 - 63(7w + 1)^2 = 1296w;$$

(5.26) *Every* $1296n$ *is of the form* $80a^2 - 17b^2 - 63c^2$, *with* $a, b, c > 0$.

A general result of the last type follows from Lemma 6, with the notation as there:

(5.27) *Every* $-pabcv^2$ *is of the form*

$$a(aw + \alpha v)^2 + b(bw + \beta v)^2 + pc(cw + \gamma v)^2.$$

An example of Lemma 2 with $c \neq -1$ is

$$(5.28)\quad 17(7w + 107v)^3 + (w - 31v)^3 - (18w + 275v)^3 = 17v^3(378w - 55v).$$

The substitution $w = 8x + 55y$, $v = -55x + 378y$ transforms (5.28) into

$$(14981x - 104940y)^3 + (1713x - 11663y)^3 - 17(5829x - 40831y)^3 = 17x(55x - 378y)^2;$$

hence *every* $17(378n - 55)^2$ *is of the form* $a^3 + b^3 - 17c^3$, *and if* $n > 7$, *all of* a, b, c *may be chosen* > 0.

6. Second method. This can be applied to any number of terms, here illustrated for 3. Let $a, \cdots, \gamma$ be such that

$$(6.1)\qquad a + b + c = \alpha + \beta + \gamma = 0, \qquad abc \neq 0.$$

Then, identically in x, y,

$$(6.2)\qquad (ax + \alpha y) + (bx + \beta y) + (cx + \gamma y) \equiv 0.$$

Two integrations of (6.2) with respect to x between the limits 0, x give

$$(6.3)\quad b^2c^2(ax+\alpha y)^3+c^2a^2(bx+\beta y)^3+a^2b^2(cx+\gamma y)^3$$
$$\equiv y^2[3abc(bc\alpha^2+ca\beta^2+ab\gamma^2)x+(b^2c^2\alpha^3+c^2a^2\beta^3+a^2b^2\gamma^3)y]$$

and it is clear that the restriction $abc\neq 0$ in (6.1) may be suppressed. A simple reduction by (6.1) gives

$$bc\alpha^2+ca\beta^2+ab\gamma^2=-(b\alpha-a\beta)^2,$$
$$b^2c^2\alpha^3+c^2a^2\beta^3+a^2b^2\gamma^3=(b\alpha-a\beta)^2[a(a+2b)\beta+b(2a+b)\alpha],$$

and we have

LEMMA 9. *Identically in x, y,*

$$b^2c^2(ax+\alpha y)^3+c^2a^2(bx+\beta y)^3+a^2b^2(cx+\gamma y)^3$$
$$\equiv-(b\alpha-a\beta)^2y^2[3abcx-\{b(2a+b)\alpha+a(a+2b)\beta\}y],$$

where $a,\cdots,\gamma$ are such that

$$a+b+c=\alpha+\beta+\gamma=0.$$

If $(a,b)=d$, and $(a,b,c)=d(a_1,b_1,c_1)$, the identity resulting from the last has (a,b,c,x) replaced by (a_1,b_1,c_1,dx), or, dropping suffixes, we recover the preceding identity with $(a,b,c)=1$ and x replaced by dx. Hence there is no loss in generality in assuming $(a,b)=1$ in Lemma 9. We can therefore choose f, g as in (4.2), and get as the solution of $b\alpha-a\beta=u$,

$$(6.4)\qquad \alpha=fu+av,\qquad \beta=gu+bv.$$

Hence, from Lemma 9, follows

LEMMA 10. *If, without loss of generality, $(a,b)=1$, and f, g are such that $bf-ag=1$, then, identically in w, z,*

$$(a+b)^2[b^2(aw+fz)^3+a^2(bw+gz)^3]-a^2b^2[(a+b)w+(f+g)z]^3$$
$$\equiv z^2[3ab(a+b)w+\{(a+2b)ag+(2a+b)bf\}z].$$

In obtaining this, the following change of notation was made, $x+vy=w$, $uy=z$. We give a few of the simplest examples. For

$$(a,b,f,g)=(3,-2,1,-1)$$

we get

$$(6.4)\quad 9(2w+z)^3+36w^3-4(3w+z)^3=z^2(18w+5z);$$

(6.5) *Every* $18n+5$ *is of the form* $9a^3+36b^3-4c^3$, *with all of* $a, b, c>0$ *if* $n>0$.

From $(a,b,f,g)=(4,-3,1,-1)$,

(6.6) $16(3w+1)^3+144w^3-9(4w+1)^3=36w+7$;

(6.7) *Every* $36n+7$ *is of the form* $16a^3+144b^3-9c^3$, *with all of* $a, b, c>0$ *if* $n>0$.

(6.8) *If* $(a,b)=1$, $bf-ag=1$, *every*

$$3ab(a+b)n+\{(a+2b)ag+(2a+b)bf\}$$

is of the form

$$a^2(a+b)^2A^3+b^2(a+b)^2B^3-a^2b^2C^3,$$

and with at most 3 *exceptions* n, *all of* A, B, C *may be chosen* >0.

Differentiating the identity in Lemma 10 with respect to w we get

(6.9) *Every* n^2 *is of the form*

$$(a+b)(aA^2+bB^2)-abC^2,$$

where $(a,b)=1$, *and all the integers* A, B, C *may be chosen* $\neq 0$ *in an infinity of ways.*

We have

$$A=bm+gn, \qquad B=am+fn, \qquad C=(a+b)m+(f+g)n,$$

where m is an arbitrary integer, and f, g are as in Lemma 10.

7. Simultaneous solutions of (2.1). Let $(a,b,c)(\alpha,\beta,\gamma)$ be two distinct solutions of (2.1). Then, identically in x, y,

$$(ax+\alpha y)^3+(bx+\beta y)^3+p(cx+\gamma y)^3\equiv 3xy(Px+Qy),$$
$$P\equiv a^2\alpha+b^2\beta+pc^2\gamma, \qquad Q\equiv a\alpha^2+b\beta^2+pc\gamma^2.$$

Write

(7.1) $A\equiv a^3+2b^3$, $B\equiv 2a^3+b^3$, $C\equiv a^3-b^3$, $D\equiv a^6+a^3b^3+b^6$,

so that $(\alpha,\beta,\gamma)=(aA,-bB,cC)$ is the first iterate of (a,b,c) obtained by (1.2). A short reduction gives

(7.2) $P=0, \qquad Q=-9pa^3b^3c^3$;

(7.3) $a^3(x+Ay)^3+b^3(x-By)^3+pc^3(x+Cy)^3=-27pa^3b^3c^3xy^2$.

(7.4) *If (a, b, c) is any solution of (2.1), every $-27pa^3b^3c^3n$, and every $-27pa^3b^3c^3n^2$, is of the form $a^3e^3 + b^3f^3 + pc^3g^3$, and with at most 3 exceptions n, in each case, $efg \neq 0$.*

In (7.3) we take $x = y$ and find

(7.5) *A two-fold infinity of solutions of*

$$x^3 + y^3 = (u^3 + v^3)(z^3 + w^3)$$

in integers x, y, z, w, u, v is

$$\begin{aligned} x &= a(1 + a^3 + 2b^3), & u &= a, & z &= 1 + a^3 - b^3, \\ y &= b(1 - 2a^3 - b^3), & v &= b, & w &= 3ab, \end{aligned}$$

where a, b are arbitrary integers.

Integrating (7.3) with respect to x from 0 to x, and reducing the constant of integration, we find (see (7.1)),

(7.6) $a^3A^4 + b^3B^4 + pc^3C^4 = 27pa^3b^3c^3D$;

(7.7) $$a^3(x + Ay)^4 + b^3(x - By)^4 + pc^3(x + Cy)^4 + 27pa^3b^3c^3Dy^4 = -54a^3b^3c^3x^2y^2;$$

(7.8) *If (a, b, c) is any solution of (2.1), every $54pa^3b^3c^3n^2$ is of the form*

$$27pa^3b^3c^3D\delta^4 - a^3\alpha^4 - b^3\beta^4 - c^3\gamma^4, \quad \textit{with} \quad \delta, \alpha, \beta, \gamma > 0$$

if n is different from $-a^3 - 2b^3$, $2a^3 + b^3$, $b^3 - a^3$.

Integrating (7.7) with respect to y from 0 to y, and reducing as before, we find

(7.9) $a^3BC - b^3CA + pc^3AB = 9pa^3b^3c^3$;

(7.10) $$\begin{aligned} a^3BC(x + Ay)^5 - b^3CA(x - By)^5 + pc^3AB(x + Cy)^5 \\ + 27pa^3b^3c^3ABCDy^5 - 9pa^3b^3c^3x^5 \\ = -90pa^3b^3c^3ABCx^2y^3; \end{aligned}$$

(7.11) *Every $-90pa^3b^3c^3ABCn^2$, and every $90pa^3b^3c^3ABCn^3$, is of the form $a^3BC\alpha^5 - b^3CA\beta^5 + pc^3AB\gamma^5 + 9pa^3b^3c^3(3ABCD\delta^5 - \epsilon^5)$, the notation being as in (2.1), (7.1), with $\alpha\beta\gamma\delta\epsilon \neq 0$ if $n \neq -A, B, -C$ for the first, and $\alpha\beta\gamma\delta\epsilon \neq 0$ for the second.*

Integration of (7.10) with respect to x between 0 and x, and reduction of the constant of integration, gives

(7.12) $a^3A^5 - b^3B^5 + pc^3C^5 = 9pa^3b^3c^3ABC$;

(7.13) $$-a^3BC(x+Ay)^6 + b^3CA(x-By)^6 - pc^3AB(x+Cy)^6 + 3pa^3b^3c^3(3x^3 - ABCy^3)(x^3 - 3ABCy^3) = 162pa^3b^3c^3ABCDxy^5;$$

(7.14) *Every* $162pa^3b^3c^3ABCDn$ *is of the form*

$$3pa^3b^3c^3(3\delta^3 - ABC\epsilon^3)(\delta^3 - 3ABC\epsilon^3) - a^3BC\alpha^6 + b^3CA\beta^6 - pc^3AB\gamma^6,$$

the notation being as in (2.1), (7.1), *and with at most* 3 *exceptions* n, *all of* $\alpha, \cdots, \epsilon$ *may be chosen* > 0.

The processes of this section can obviously be continued indefinitely. From (7.12) we note that

(7.15) *For an infinity of integers* p,

$$x^3u^5 + y^3v^5 - pz^3w^5 = 9px^3y^3z^3uvw$$

is solvable in integers $x, \cdots$ *with* $xyzuvw \neq 0$.

California Institute of Technology,
Pasadena, California.

ON THE NORMAL FORMS OF LINEAR CANONICAL TRANSFORMATIONS IN DYNAMICS.

By JOHN WILLIAMSON.

Let n be the number of degrees of freedom of a linear conservative dynamical system and let the point $(q_1, q_2, \cdots, q_n;\ p_1, p_2, \cdots, p_n)$ of the phase space be denoted by $x = (x_1, x_2, \cdots, x_{2n})$. A system of $2n$ ordinary differential equations of the first order, which are homogeneous, linear and do not contain t explicitly, is a canonical system, if, and only if, the differential equations can be written in the form

$$G\frac{dx}{dt} = Hx,$$

where H is a real symmetric matrix of order $2n$ and G is the skew symmetric matrix $\begin{pmatrix} 0 & E \\ -E & 0 \end{pmatrix}$, and E the unit matrix of order n. A non-singular linear transformation

$$y = Ax$$

is said to be a canonical transformation, if it transforms every linear canonical system into a linear canonical system. It is known that the transformation of matrix A is canonical if, and only if,

$$\text{(i)} \qquad A'GA = sG$$

where s is a constant.[1] It can be assumed without loss of generality that[2] $s = 1$ and accordingly we shall call a matrix A a canonical matrix if it satisfies (i) with $s = +1$.

In a previous paper[3] normal forms for dynamical systems under canonical transformations were found and here we determine normal forms for canonical matrices under canonical transformations. These normal forms are not completely determined by the elementary divisors of the canonical matrix, so that two canonical matrices, which are similar, are not necessarily similar under a canonical transformation.

---

[1] A. Wintner, "On the linear conservative dynamical systems," *Annali di matematica pura ed applicata*, ser. 4, tomo 13 (1934-35), pp. 105-112.

[2] E. R. van Kampen and A. Wintner, "On the canonical transformations of Hamiltonian systems," *American Journal of Mathematics*, vol. 58 (1936), pp. 851-863.

[3] John Williamson, "On the algebraic problem concerning the normal forms of linear dynamical systems," *American Journal of Mathematics*, vol. 58 (1936), pp. 141-163.

When considering the solutions of the equations of variation belonging to a periodic solution of conservative non-linear dynamical systems, the question of the occurrence of secular terms is known to depend on the elementary divisors of a canonical matrix.[4] In fact the degree of the highest secular term occurring is determined by the greatest exponent of the elementary divisors. For this reason, it is of interest, that it is possible for a canonical matrix to have an elementary divisor of order $2m$ (§ 6, Result III_a).[5]

In the following sections the problem is considered from a purely algebraic point of view and in section 1 is reduced to a simpler one of a similar nature; sections 2 and 3 are devoted to the proofs of preliminary lemmas, while the main results are obtained in the remaining sections.

1. Simplification of the problem. Let E be the unit matrix of order n and G the skew-symmetric matrix $G=\begin{pmatrix} 0 & E \\ -E & 0 \end{pmatrix}$ of order $2n$. A real matrix A_i is said to be a *canonical matrix*, if

$$(1) \qquad A_iGA'_i = G,$$

where A'_i is the transposed of A_i. We shall be interested in determining necessary and sufficient conditions that two canonical matrices A_1 and A_2 be similar under a *canonical transformation*; in other words that there exist a third canonical matrix A_3, such that

$$(2) \qquad A_2 = A_3A_1A_3^{-1}.$$

We first reduce this problem to a somewhat simpler one.

If A_1 and A_2 are two canonical matrices, which are similar, and a matrix Q, to be specified later, is similar to A_1, then Q is similar to A_2. There exist, therefore, two non-singular matrices R_1 and R_2, such that

$$(3) \qquad R_iA_iR_i^{-1} = Q \qquad (i=1,2).$$

The matrices

$$(4) \qquad R_iGR'_i = S_i \qquad (i=1,2),$$

are skew symmetric and are left invariant by Q, that is, satisfy the equations,

---

[4] A. Wintner, "Three notes on characteristic exponents and equations of variation in celestial mechanics," *American Journal of Mathematics*, vol. 53 (1931), pp. 605-625.

[5] This result could be deduced by suitable modifications from papers by Alfred Loewy, "Allgemeine bilineare Formen mit konjugirt imaginären Variabeln," *Abhandlungen der Kaiserlichen Leopoldinisch-Carolinischen Deutschen Akademie der Naturforscher*, Band 71. S. S. 378-446, Halle (1898), and T. J. I'A. Bromwich, "Canonical reduction of bilinear forms," *Proceedings of the London Mathematical Society*, vol. 32 (1900), pp. 321-332.

$$QS_iQ' = S_i \qquad (i = 1, 2). \tag{5}$$

Thus, if Q is any matrix similar to the two canonical matrices A_1 and A_2, there is associated with A_1 a skew symmetric matrix S_1 and with A_2 a skew symmetric matrix S_2, both of which are left invariant by Q.

We now prove

THEOREM 1. *A necessary and sufficient condition, that A_1 be similar to A_2 under a canonical transformation, is that there exist a non-singular matrix H, such that*

$$HQ = QH, \tag{6}$$

and that the two skew symmetric matrices S_1 and S_2, associated with A_1 and A_2, satisfy

$$HS_1H' = S_2. \tag{7}$$

Proof. Let a matrix H satisfying (6) and (7) exist. Then

$$\begin{aligned} A_2 &= R_2^{-1}QR_2 = R_2^{-1}HQH^{-1}R_2 \text{ by (3) and (6)}, \\ &= R_2^{-1}HR_1A_1R_1^{-1}H^{-1}R_2 \text{ by (3)}, \\ &= A_3A_1A_3^{-1}, \end{aligned}$$

where $A_3 = R_2^{-1}HR_1$. Further

$$\begin{aligned} A_3GA'_3 &= R_2^{-1}HR_1GR'_1H'(R_2^{-1})' = R_2^{-1}HS_1H'(R_2^{-1})' \text{ by (4)}, \\ &= R_2^{-1}S_2(R_2^{-1})' = G \text{ by (7) and (4)}. \end{aligned}$$

Hence A_3 is a canonical matrix. Conversely, if (2) is satisfied and A_3 is a canonical matrix, the matrix $H = R_2A_3R_1^{-1}$ satisfies (6) and (7); for

$$HQH^{-1} = R_2A_3R_1^{-1}QR_1A_3^{-1}R_2^{-1} = R_2A_3A_1A_3^{-1}R_2^{-1} = R_2A_2R_2^{-1} = Q \quad \text{and}$$
$$HS_1H' = R_2A_3R_1^{-1}S_1(R_1^{-1})'A'_3R'_2 = R_2A_3GA'_3R'_2 = R_2GR'_2 = S_2.$$

Since, in the above, Q is any matrix similar to A_1, we are at liberty to choose Q in a suitable normal form. Then, if S is any real skew symmetric matrix satisfying the equation

$$QSQ' = S, \tag{8}$$

we shall determine a normal form for S under transformations by matrices permutable with Q. If $HQ = QH$ and $HSH' = S_1$, we shall call the transformation by the matrix H an *admissible transformation* and shall write $S \approx S_1$.

2. Preliminary lemmas. When R is a square matrix of order m, we may consider R as a matrix of matrices and write

$$(9) \qquad R = (R_{ij}) \qquad (i, j = 1, 2, \cdots, t),$$

where R_{ij} is a matrix of r_i rows and r_j columns and $r_1 + r_2 + \cdots + r_t = m$. If S is a second m-rowed square matrix and S is written as a matrix of matrices

$$(10) \qquad S = (S_{ij}) \qquad (i, j = 1, 2, \cdots, t),$$

where S_{ij} is also a matrix of r_i rows and r_j columns, we shall say that R and S are *similarly partitioned* or that (10) is a partition of S similar to that of R in (9). If in (9), when i is different from j, R_{ij} is the zero matrix, we shall call R a *diagonal block* matrix and write

$$R = [R_{11}, R_{22}, \cdots, R_{tt}].$$

LEMMA 1. *If the matrices S_1, S_2 and Q satisfy (5), then $S_2 = MS_1$, where $MQ = QM$.*

Proof. Since S_1 and S_2 are non-singular, Q is non-singular and accordingly

$$(Q')^{-1} = S_1^{-1}QS_1 = S_2^{-1}QS_2,$$

so that

$$S_2S_1^{-1}Q = QS_2S_1^{-1}.$$

If $M = S_2S_1^{-1}$, then $MQ = QM$ and $S_2 = MS_1$.

LEMMA 2. *If $Q = [Q_1, Q_2]$ and no latent root of Q_1 is the reciprocal of a latent root of Q_2, a matrix S, which satisfies (8), is of the form $[S_{11}, S_{22}]$ and*

$$Q_iS_{ii}Q'_i = S_{ii} \qquad (i = 1, 2).$$

Proof. Let

$$S = (S_{ij}) \qquad (i, j = 1, 2),$$

be a partition of S similar to that of Q. Then

$$(11) \qquad Q_iS_{ij} = S_{ij}(Q'_j)^{-1} \qquad (i, j = 1, 2).$$

Since, by hypothesis, no latent root of Q_1 is the reciprocal of a latent root of Q_2, no latent root of Q_1 is the same as a latent root of $(Q'_2)^{-1}$. Therefore, as a consequence of (11), $S_{12} = 0$. Similarly $S_{21} = 0$ and the lemma is proved.

LEMMA 3. *Let $Q = [Q_1, Q_2]$ and let S be a skew-symmetric matrix satisfying (8). If $S = (S_{ij})$ $(i, j = 1, 2)$, is a partition of S similar to that of Q and, if S_{11} is non-singular, then*

$$S \approx S_1,$$

where $S_1 = [S_{11}, T_{22}]$.

Proof. Let E_i be the unit matrix of the same order as Q_i and H be the matrix

$$H = \begin{pmatrix} E_1 & 0 \\ -S_{21}S_{11}^{-1} & E_2 \end{pmatrix}.$$

As a consequence of (11),

$$S_{21}S_{11}^{-1}Q_1 = S_{21}(Q'_1)^{-1}S_{11}^{-1} = Q_2S_{21}S_{11}^{-1}.$$

Hence $HQ = QH$. Since S is skew-symmetric,

$$HSH' = \begin{pmatrix} E_1 & 0 \\ -S_{21}S_{11}^{-1} & E_2 \end{pmatrix}\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}\begin{pmatrix} E_1 & -S_{11}^{-1}S_{12} \\ 0 & E_2 \end{pmatrix} = [S_{11}, T_{22}],$$

where $T_{22} = S_{22} - S_{21}S_{11}^{-1}S_{12}$.

3. Normal form of Q. Let p be a real number or else the two rowed real matrix

$$p = \begin{pmatrix} a & b \\ -b & a \end{pmatrix}, \text{ where } b \neq 0.$$

Let E_i denote the unit matrix of order e_i and U_i the auxiliary unit matrix of the same order.[6] The matrix,

$$P_i = pE_i + pU_i, \tag{12}$$

has the single elementary divisor $(\lambda - p)^{e_i}$ or the two elementary divisors $(\lambda - a + ib)^{e_i}$, $(\lambda - a - ib)^{e_i}$, according as p is a matrix of order one or two.[7] The diagonal block matrix,

$$\pi = [P_1, P_2, \cdots, P_t], \tag{13}$$

has therefore the elementary divisors $(\lambda - p)^{e_j}$ or $(\lambda - a + ib)^{e_j}, (\lambda - a - ib)^{e_j}$ $(j = 1, 2, \cdots, t)$. We may take Q in the normal form,

$$Q = [\pi_1, \pi_2, \cdots, \pi_k], \tag{14}$$

where the matrix π_j is obtained from π in (13) by writing p_j for p, e_{ij} for e_i, and t_j for t. Further $p_j \neq p_i$, if j is different from i.

If H is a matrix commutative with Q, H is a diagonal block matrix $[H_1, H_2, \cdots, H_k]$, where

$$H_j\pi_j = \pi_jH_j \qquad (j = 1, 2, \cdots, k). \tag{15}$$

---

[6] Cf. Turnbull and Aitken, *Canonical Matrices*, p. 62.

[7] By the elementary divisors of a matrix A we mean the elementary divisors of $A - \lambda E$.

[8] John Williamson, "The idempotent and nilpotent elements of a matrix," *American Journal of Mathematics*, vol. 58 (1936), p. 477.

But the form of a matrix H_j satisfying (15) is known.[8] In fact, if $W\pi = \pi W$ and $W = (W_{ij})$ $(i, j = 1, 2, \cdots, t)$, is a partition of W similar to that of π in (13) and, if $e_i \geqq e_j$, then

$$W_{ij} = \begin{pmatrix} F_{ij} \\ 0 \end{pmatrix} \quad \text{and} \quad W_{ji} = (0, F_{ji}),$$

where F_{ij} and F_{ji} are square matrices of order e_j. Moreover F_{ij} and F_{ji} are both polynomials in U_j with coefficients, which are polynomials in p. More exactly

$$F_{ij} = \sum_{a=0}^{e_j-1} f_{ija}(p) U_j{}^a,$$

while

$$F'_{ij} = \sum_{a=0}^{e_j-1} f_{ija}(p') U'_j{}^a. \tag{16}$$

If i denotes the matrix $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ and p is a two rowed matrix,

$$p = a + ib \quad \text{and} \quad p' = a - ib = \bar{p}.$$

With this notation (16) becomes

$$F'_{ij} = \sum_{a=0}^{e_j-1} \bar{f}_{ija} U'_j{}^a. \tag{17}$$

Let T_j be the counter unit matrix of order e_j. Then

$$T_j U_j = U'_j T_j \tag{18}$$

and as a consequence of (17)

$$T_j F'_{ij} = \bar{F}_{ij} T_j. \tag{19}$$

LEMMA 4. *Let $T_j W'_{ij} = \tilde{W}_{ij} T_i$. If $e_i = e_j$, $\tilde{W}_{ij} = \bar{W}_{ij}$. If $e_i > e_j$, the element in the first row and first column of $\tilde{W}_{ij}$ is zero.*

Proof. Let $e_i \geqq e_j$. Then

$$\begin{aligned} T_j W'_{ij} &= (T_j F'_{ij}, 0) = (\bar{F}_{ij} T_j, 0) \quad \text{by (19)}, \\ &= (0, \bar{F}_{ij}) T_i. \end{aligned}$$

Hence $\tilde{W}_{ij} = (0, \bar{F}_{ij})$ and the lemma is proved.

4. Reduction of S. Let

$$Q = [Q_1, Q_2, \cdots, Q_k], \tag{20}$$

where no latent root of Q_1 has absolute value 1, each latent of Q_2 is equal to 1, each latent root of Q_3 is equal to -1 and each latent roof of Q_j, $j > 3$, is equal to $a_j \pm ib_j$, where $a_j{}^2 + b_j{}^2 = 1$ and $a_j + ib_j \neq a_r + ib_r$ unless $r = j$. Then,

if S is a skew-symmetric matrix satisfying (8), as a consequence of Lemma 2, $S=[S_1, S_2, \cdots, S_k]$, where

$$(21) \qquad Q_j S_j Q'_j = S_j, \qquad (j=1,2,\cdots,k).$$

Since any matrix H, commutative with the matrix Q in (20), is also a diagonal block matrix, we may consider each of the equations (21) separately. Since S_1 is non-singular, Q_1 is similar to $(Q'_1)^{-1}$ and, since no latent root of Q_1 has absolute value 1, Q_1 is similar to a matrix $[F_1, (F'_1)^{-1}]$, where the order of F_1 is one-half that of Q_1. Hence Q_1 may be replaced by the matrix $[F_1, (F'_1)^{-1}]$. It is now a consequence of (21) that

$$S_1 = \begin{pmatrix} 0 & \sigma \\ -\sigma & 0 \end{pmatrix}, \text{ where } \sigma F_1 = F_1 \sigma.$$

If $H_1 = \begin{pmatrix} \sigma^{-1} & 0 \\ 0 & I_1 \end{pmatrix}$, where I_1 is the unit matrix of the same order as F_1,

$$H_1[F_1, (F'_1)^{-1}]H_1^{-1} = [F_1, (F'_1)^{-1}] \text{ and } H_1 S_1 H'_1 = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} = G_1.$$

Hence we have

Result I. The matrix Q_1 may be taken in the form $Z_1 = [F_1, (F'_1)^{-1}]$. With this value of Q_1, $S_1 \approx G_1$.

The matrix F_1 is not unique and may be replaced by any matrix similar to it; in fact F_1 may be taken in the normal form $[\pi_1, \pi_2, \cdots, \pi_f]$, where π_j is defined by (13) and $|p_j| \neq 1$. As a consequence of the above and Theorem 1 we have

THEOREM 2. *If A_1 is a canonical matrix similar to a second canonical matrix A_2 and, if no latent root of A_1 is of absolute value* 1, *A_1 is similar to A_2 under a canonical transformation.*

We next consider equations (21), when $j \geqq 2$, and for simplicity of notation temporarily drop the suffix j. The matrix $Q = Q_j$ is therefore of the form

$$(22) \qquad Q = [P_1, P_2, \cdots, P_t], \qquad e_1 \geqq e_2 \geqq \cdots \geqq e_t,$$

where P_i is defined by (12) with the added restriction that $|p|=1$. Hence p is a real orthogonal matrix of order one or two. If

$$S = (S_{ij}), \qquad (i, j = 1, 2, \cdots, t),$$

is a partition of S similar to that of Q in (22), equation (18) implies

$$P_i S_{ij} P'_j = S_{ij}, \qquad (i, j = 1, 2, \cdots, t),$$

or, if $S_{ij} = \sigma$,

$$(23) \qquad P_i \sigma P'_j = \sigma.$$

The matrix $\sigma = (\sigma_{rs})$ in (23) is a matrix of $m = e_i$ rows and $n = e_j$ columns. On equating corresponding elements in (23) we obtain

(24) $p(\sigma_{rs} + \sigma_{r+1,s} + \sigma_{r,s+1} + \sigma_{r+1,s+1})p' = \sigma_{rs}$, $(r = 1, 2, \cdots, m; s = 1, 2, \cdots, n)$,

with the understanding that $\sigma_{m+1,s} = \sigma_{r,n+1} = 0$. If $p = \pm 1$, (24) reduces to

$$\sigma_{r+1,s} + \sigma_{r,s+1} + \sigma_{r+1,s+1} = 0. \tag{25}$$

On substituting $s = n, n-1, n-2, \cdots$, successively in (25) we have

(26) $\sigma_{r+1,n} = \sigma_{r+2,n-1} = \cdots = \sigma_{r+s+1,n-s} = 0$, $(r = 1, 2, \cdots, m;\ s = 1, 2, \cdots, n)$,

and, on substituting $r = m, m-1, m-2, \cdots$, successively,

(27) $\sigma_{m,s+1} = \sigma_{m-1,s+2} = \cdots = \sigma_{m-r,r+s+1} = 0$, $(r = 1, 2, \cdots, m; s = 1, 2, \cdots, n)$.

We easily deduce from (25), (26) and (27),

LEMMA 5. *If σ is a matrix satisfying* (23) *and, if $e_i \neq e_j$, the last row and the last column of σ are zero. If $e_i = e_j = n$, then $\sigma_{rs} = 0$, when $r + s > n + 1$ and*

$$\sigma_{n1} = -\sigma_{n-1,2} = \sigma_{n-2,3} = \cdots = (-1)^{n-1}\sigma_{1n}. \tag{28}$$

If p is of order 2, equations (25) may be solved to give a particular matrix S^*, which satisfies (22) and whose elements are two rowed scalar matrices. Any other matrix S, which satisfies (22) is, by Lemma 1, of the form MS^*, where $MQ = QM$. Since the elements of M are polynomials in p, so are the elements of S. Hence

$$p\sigma_{rs}p' = \sigma_{rs}pp' = \sigma_{rs}, \tag{29}$$

since p is orthogonal. Accordingly, *Lemma 5 is also true, when p is a two-rowed matrix.*

Let $e_1 = e_2 = \cdots = e_c > e_{c+1}$ and let s_{ij} denote the element in the first column and the last row of S_{ij}. Then, by Lemma 5, S_{11} is singular, if and only if s_{11} is zero. If S_{jj} is non-singular, $1 < j \leqq c$, we may interchange S_{jj} and S_{11} without disturbing Q. If S_{jj} is singular for all values of j, $1 \leqq j \leqq c$, then

$$s_{jj} = 0, \qquad (j = 1, 2, \cdots, c). \tag{30}$$

Since S is non-singular and since, by Lemma 5, the last row of S_{1k} is zero, when $k > c$, for at least one value of j, $1 < j \leqq c$, $s_{1j} \neq 0$. We may therefore suppose, without any loss of generality, that $s_{12} \neq 0$.

Let I be the unit matrix of order $e_3 + e_4 + \cdots + e_t$ and H_1 the matrix $\left[\begin{pmatrix} E_1 & E_1 \\ 0 & E_1 \end{pmatrix}, I\right]$. Then H_1 is commutative with Q and

$$H_1SH'_1 = R = (R_{ij}) \qquad (i, j = 1, 2, \cdots, t),$$

where

$$R_{11} = S_{11} + S_{12} + S_{21} + S_{22}.$$

The element in the last row and first column of R_{11} is

$$r_{11} = s_{11} + s_{12} + s_{21} + s_{22} = s_{12} + s_{21} \text{ by (30)}.$$

If p is a two-rowed matrix the transformation by the matrix

$$H_2 = \left[\begin{pmatrix} E_1 & -iE_1 \\ 0 & E_1 \end{pmatrix}, I\right]$$

is admissible and

$$H_2SH'_2 = F = (F_{ij}), \qquad (i, j = 1, 2, \cdots, t),$$

where

$$f_{11} = i(s_{21} - s_{12}).$$

Since S is skew symmetric, $S_{21} = -S'_{12}$ and by (28)

$$s_{21} = -(-1)^{e_1-1}s'_{12}.$$

Hence, if e_1 is even and p is of order 1, $s_{21} = s'_{12} = s_{12}$ and $r_{11} = 2s_{12} \neq 0$. If p is of order 2, at least one of f_{11} or r_{11} is different from zero and accordingly at least one of F_{11} or R_{11} is non-singular. Therefore, unless e_1 is odd and $p = \pm 1$,

$$S \approx L = (L_{ij}), \qquad (i, j = 1, 2, \cdots, t),$$

where

$$L_{11} = S_1$$

is non-singular.

If e_1 is odd and $p = \pm 1$, $s_{11} = -s'_{11} = 0$ and S_{11} is singular. Hence $c \geqq 2$ and we may suppose that $s_{12} \neq 0$. Then the matrix

$$S_1 = (S_{ij}) \qquad (i, j = 1, 2),$$

is non-singular, since

$$| S_1 | = \pm (s_{12})^{2e_1},$$

as is seen by re-arranging the rows and columns of S_1 in the order 1, $e_1 + 1$, 2, $e_1 + 2$, etc. By repeated applications of Lemma 3 we therefore deduce that

$$S \approx [S_1, S_2, \cdots, S_k]. \tag{31}$$

The component matrices S_j on the right of (31) are of two distinct types:

Type a. The matrix S_j is of order $2e_j$, $p = \pm 1$, e_j is odd, and $[P_j, P_j]S_j[P_j, P_j]' = S_j$.

Type b. The matrix S_j is of order e_j and $P_jS_jP'_j = S_j$.

Reduction of type a. For convenience we drop the suffix j and write

$$[P, P]S[P, P]' = S, \text{ where } S = (S_{rs}), \qquad (r, s = 1, 2).$$

Hence

$$PS_{rs}P' = S_{rs} \qquad (r, s = 1, 2).$$

As a consequence of Lemma 1,

$$(32) \qquad S_{rs} = M_{rs}X,$$

where $M_{rs} = M_{rs}(U)$ is a polynomial in $U = U_j$ and X is a particular solution of $PXP' = X$. Since S_{rr} is singular and X is non-singular,

$$(33) \qquad M_{rr}(U) = Um_{rr}(U); \qquad (r = 1, 2).$$

If

$$\sigma_1 = \begin{pmatrix} 0 & S_{12} \\ S_{21} & 0 \end{pmatrix} \text{ and } \sigma_2 = \begin{pmatrix} S_{11} & 0 \\ 0 & S_{22} \end{pmatrix},$$

σ_1 is non-singular and

$$(34) \qquad \sigma_2\sigma_1^{-1} = \begin{pmatrix} 0 & M_{11}M_{21}^{-1} \\ M_{22}M_{12}^{-1} & 0 \end{pmatrix}.$$

The matrix $\sigma_2\sigma_1^{-1}$ is commutative with $[P, P]$ and therefore so is the matrix $H = E - \frac{1}{2}\sigma_2\sigma_1^{-1}$. Further

$$HSH' = (E - \tfrac{1}{2}\sigma_2\sigma_1^{-1})(\sigma_1 + \sigma_2)(E - \tfrac{1}{2}\sigma_1^{-1}\sigma_2),$$
$$= \tau_1 + \tau_2,$$

where $\tau_1 = \sigma_1 - \frac{3}{4}\sigma_2\sigma_1^{-1}\sigma_2$ and $\tau_2 = \frac{1}{4}\sigma_2\sigma_1^{-1}\sigma_2\sigma_1^{-1}\sigma_2$. As a consequence of (32), (33), and (34)

$$\tau_2 = [K_{11}X, K_{22}X],$$

where K_{11} and K_{22} are polynomials in U each with a factor U^3, while τ_1 is of the same nature as σ_1 and is non-singular. We may therefore repeat this process of reduction with σ_i replaced by τ_i and, since $U^{e_j} = 0$, in at most $(e_j + 1)/2$ steps reduce S to the form

$$\begin{pmatrix} 0 & S_{12} \\ -S'_{12} & 0 \end{pmatrix} \text{ where } PS_{12}P' = S_{12}.$$

Let H be the matrix $H = [E_j, (S'_{12})^{-1}]$. Then

$$HSH' = \begin{pmatrix} 0 & E_j \\ -E_j & 0 \end{pmatrix} = G_j,$$

and

$$H[P, P]H' = [P, (S'_{12})^{-1}PS'_{12}] = [P, (P')^{-1}].$$

We have therefore

Result II. In type a the matrix $[P_j, P_j]$ may be replaced by

$$Z_j = [P_j, (P'_j)^{-1}].$$

Then $S_j \approx G_j$.

Reduction of type b. Again we drop the suffix j and write $P = pE + pU$, where P is of order e. If T is the counter unit matrix of order e, it is a consequence of Lemma 5 that

$$S = (\sigma_0 + \sigma_1 + \cdots + \sigma_{e-1})T, \tag{37}$$

where the elements of σ_k are all zero except in the k-th diagonal above the leading one. If σ_{jk} is the non-zero element in the j-th row of σ_k, a simple calculation shows that

$$\sigma_{jk} = s_{k,e+1-j-k}. \tag{38}$$

The matrix $U^k\sigma_j$ is of the same type as σ_{k+j} and, in particular, the elements of

$$\rho_k = U^k\sigma_0 \tag{39}$$

are all zero except those in the k-th diagonal above the leading one. The non-zero element in the j-th row of ρ_k is

$$\rho_{jk} = \sigma_{j+k,0} = s_{j+k,e+1-j-k}. \tag{40}$$

If

$$H_k = E + qU^k, \text{ where } qp = pq,$$

then

$$H_kS_kH'_k = C = (c_{rs}), \qquad (r, s = 1, 2, \cdots, e). \tag{41}$$

Since $TU' = UT$,

$$\begin{aligned} C &= (E + qU^k)STT(E + q'U'^k) \\ &= (E + qU^k)(\sigma_0 + \sigma_1 + \cdots \sigma_{e-1})(E + q'U^k)T \\ &= (\gamma_0 + \gamma_1 + \cdots + \gamma_{e-1})T, \end{aligned}$$

where

$$\gamma_f = \sigma_f, \qquad (f = 0, 1, 2, \cdots, k-1), \tag{42}$$

and

$$\gamma_k = \sigma_k + U^kq\sigma_0 + \sigma_0q'U^k.$$

Since, $\sigma_0U = - U\sigma_0$, this last equation becomes

$$\gamma_k = \sigma_k + (q + (-1)^kq')\sigma_0U^k.$$

The non-zero element in the j-th row of γ_k is (cf. 38)

$$c_{j,e+1-j-k} = \sigma_{jk} + (q + (-1)^kq')\rho_{jk}.$$

Hence by (38) and (40),

$$(43) \qquad c_{j,e+1-j-k} = s_{j,e+1-j-k} + (q + (-1)^k q') s_{j+k,e+1-j-k}.$$

Type b_1. $e = 2m$. Let $k = e + 1 - 2j$ and $q = -s_{jj}(2s_{e+1-j,j})^{-1}$. Then, since s_{jj} is skew symmetric and $s_{e+1-j,j}$ is symmetric, q is skew symmetric and as a consequence of (43), $c_{jj} = 0$.

Since, by (25), $s_{j,j-1} + s_{j-1,j} + s_{jj} = 0$, if $s_{jj} = 0$, $s_{j,j-1} = -s_{j-1,j}$ and accordingly $s_{j,j-1}$ is symmetric. Therefore, when $s_{jj} = 0$, if $k = e + 2 - 2j$ and $q = -s_{j,j-1}(2s_{e+2-j,j-1})^{-1}$, it is a consequence of (43) that $c_{j,j-1} = 0$. Hence it is possible by an admissible transformation to reduce S to a form, in which $s_{jj} = s_{j,j-1} = s_{j-1,j} = 0$. Equations (42) show that such a transformation does not alter the value of s_{rs}, when $r + s > 2j$. Therefore by giving j successively the values $m, m-1, \cdots, 2, 1$ we deduce that $S \approx D$, where

$$(44) \qquad d_{11} = d_{12} = d_{21} = d_{22} = d_{23} = \cdots = d_{m,m-1} = d_{m,m} = 0.$$

Equations (44) and (25) together imply that

$$(45) \qquad d_{rs} = 0, \qquad (r, s = 1, 2, \cdots, m).$$

The non-zero elements of D are now determined by means of equations (25) in the form

$$d_{rs} = d_{1e} x_{rs},$$

where x_{rs} is unique. Hence

$$S = dX$$

where X is uniquely determined. Since $d = d_{1e}$ is symmetric, d is a scalar. Therefore the admissible transformation of matrix $E/\sqrt{d}$ reduces dX to the form ϵX, where $\epsilon = \pm 1$, so that

$$S \approx \epsilon X, \qquad \epsilon = \pm 1.$$

As a consequence of (45), we have

$$(46) \qquad X = \begin{pmatrix} 0 & X_{12} \\ -X'_{12} & 0 \end{pmatrix},$$

where X_{12} is a square matrix of order $m = e/2$. For example if $m = 4$,

$$X_{12} = \begin{pmatrix} 1 & 3 & 3 & 1 \\ -1 & -2 & -1 & 0 \\ 1 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix},$$

where, in case p is a two-rowed matrix, each integer denotes the corresponding scalar matrix.[9] We have therefore

Result III. In type b, when $e_j = 2m_j$, $S_j \approx \epsilon X_j$, where $\epsilon = \pm 1$ and X_j is uniquely determined.

Type b_2. $e = 2m + 1$. The matrix p is necessarily a two-rowed matrix and each element s_{rs} of S is of the form $s_{rs} = a_{rs} + ib_{rs}$. Further, since

$$s_{e1} = (-1)^{2m} s_{1e} = -s'_{e1},$$

s_{e1} is skew symmetric and, as a consequence of (28), s_{rs} is skew symmetric, when $r + s = e + 1$. If $k = 1$, and $q = -a_{m,m+1}(2s_{m+1,m+1})^{-1}$, as a consequence of (43),

$$c_{m,m+1} = s_{m,m+1} - a_{m,m+1} = ib_{m,m+1}.$$

Therefore we may suppose $s_{m,m+1}$ to be skew symmetric. By a process analogous to that adopted for the case $e = 2m$ it may be shown that

$$S \approx \epsilon Y,$$

where $\epsilon = \pm 1$ and Y is uniquely determined. In particular

$$(47) \qquad y_{rs} = 0, \qquad (r, s = 1, 2, \cdots, m),$$

and

$$(48) \qquad y_{rs} = 0, \qquad r + s > e + 2.$$

For example, if $e = 5$,

$$Y = \begin{pmatrix} 0 & 0 & i/2 & 3i/2 & i \\ 0 & 0 & -i/2 & -i & 0 \\ i/2 & -i/2 & i & 0 & 0 \\ 3i/2 & -i & 0 & 0 & 0 \\ i & 0 & 0 & 0 & 0 \end{pmatrix}.$$

We have therefore

Result IV. In type b_2 when $e_j = 2m_j + 1$, $S_j \approx \epsilon Y_j$ where $\epsilon = \pm 1$ and Y_j is uniquely determined.

By combining results I, II, III, and IV it is possible to determine a normal form for S under admissible transformations. This normal form is not completely determined by the elementary divisors of $Q - \lambda E$ or of $A - \lambda E$. *With each elementary divisor of the form $(\lambda \pm 1)^{2k}$ and with each pair of conjugate elementary divisors of the form $(\lambda - a \pm ib)^k$, $a^2 + b^2 = 1$, is associated a positive or negative sign.*

---

[9] Cf. Turnbull and Aitken, *Canonical Matrices*, pp. 155-159.

Before proceeding to show that the elementary divisors together with the signs attached to them completely determine the normal form for S, we deduce

THEOREM 3. *If A is a canonical matrix the determinant of A has the value* 1.

This is an immediate consequence of the fact that the determinant of A is the product of the latent roots of A and that the latent root -1 must occur an even number of times (result II).

5. Necessary conditions. Let A_1 and A_2 be two canonical matrices, which are similar under a canonical transformation. Then, by Theorem I, the associated skew-symmetric matrices S_1 and S_2 are equivalent under an admissible transformation. The matrices S_1 and S_2 may be taken in the normal form of the previous section and are accordingly diagonal block matrices, whose component block matrices differ at most in sign. If, with the notation of (20),

$Q = [Q_1, Q_2, \cdots, Q_k]$, $\quad S_1 = [\sigma_1, \sigma_2, \cdots, \sigma_k]$ and $\quad S_2 = [\tau_1, \tau_2, \cdots, \tau_k]$,

there then exist k non-singular matrices W_j such that

$$W_j \sigma_j W'_j = \tau_j \text{ and } W_j Q_j = Q_j W_j, \qquad (j = 1, 2, \cdots, k).$$

We need, therefore, consider only equations of the type

$$W \sigma W' = \tau, \qquad W \pi = \pi W,$$

where π is defined by (13). If $\sigma = (\sigma_{ij})$, $\tau = (\tau_{ij})$ and $W = (W_{ij})$, $(i, j = 1, 2, \cdots, t)$, we have the equations

$$\sum_{\alpha=1}^{t} \sum_{\beta=1}^{t} W_{i\alpha} \sigma_{\alpha\beta} W'_{j\beta} = \tau_{ij}, \qquad (i, j = 1, 2, \cdots, t). \tag{49}$$

If $\sigma_{ij} = K_{ij} T_j$ and $\tau_{ij} = F_{ij} T_j$, equation (49) becomes

$$\sum_{\alpha=1}^{t} \sum_{\beta=1}^{t} W_{i\alpha} K_{\alpha\beta} T_\beta \bar{W}'_{j\beta} = F_{ij} T_j,$$

or, by Lemma 4,

$$\sum_{\alpha=1}^{t} \sum_{\beta=1}^{t} W_{i\alpha} K_{\alpha\beta} \bar{W}_{j\beta} T_j = F_{ij} T_j \qquad (i, j = 1, 2, \cdots, t).$$

It is a consequence of the nature of the matrices W_{ij}, K_{ij}, etc., that this last equation implies

$$\sum_{\alpha=1}^{t} \sum_{\beta=1}^{t} w_{i\alpha} k_{\alpha\beta} \bar{w}_{j\beta} = f_{ij}, \qquad (i, j = 1, 2, \cdots, t), \tag{50}$$

where each small letter denotes the element in the first row and the first column of the matrix denoted by the corresponding capital letter. Since σ and τ are in normal form $\sigma_{ij} = \tau_{ij} = 0$, if $e_i \neq e_j$. Further $w_{ij} = 0$, if $e_i < e_j$, and, by Lemma 4, $\bar{w}_{ij} = 0$, if $e_i > e_j$. Hence, if $e_{c-1} > e_c = e_{c+1} = \cdots = e_d > e_{d+1}$, we have as a result of (50) and Lemma 4,

$$(51) \qquad \sum_{\alpha=c}^{d} \sum_{\beta=c}^{d} w_{i\alpha} k_{\alpha\beta} \bar{w}_{j\beta} = f_{ij}, \qquad (i, j = c, c+1, \cdots, d).$$

If B is the matrix whose elements are w_{ij}, $(i, j = c, c+1, \cdots, d)$, (50) may be written in the form

$$(52) \qquad B(k_{ij})\bar{B}' = (f_{ij}), \qquad (i, j = c, c+1, \cdots, d).$$

Since $|B|$ is a factor of $|W|$ and W is non-singular so is B.[10] The $(d-c)$-rowed square matrices (k_{ij}) and (f_{ij}), $(i, j = c, c+1, \cdots, d)$, are therefore conjunctively equivalent. Further, as a consequence of results I-IV, (f_{ij}) coincides with (k_{ij}), unless the matrix P_i is of type b. In this last case

$$(k_{ij}) = [\epsilon_c g, \epsilon_{c+1} g, \cdots, \epsilon_d g] \quad \text{and} \quad (f_{ij}) = [\epsilon'_c g, \epsilon'_{c+1} g, \cdots, \epsilon'_d g],$$

where $\epsilon_j = \pm 1$, $\epsilon'_j = \pm 1$ and $g = 1$ or $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Therefore we deduce from (52) that

$$B[\epsilon_c, \epsilon_{c+1}, \cdots, \epsilon_d]\bar{B}' = [\epsilon'_c, \epsilon'_{c+1}, \cdots, \epsilon'_d].$$

Hence the number of positive ϵ_j is the same as the number of positive ϵ'_j. We may call the number of positive ϵ_j *the index of the elementary divisors* $(\lambda \pm 1)^{e_c}$ *or of the pair of conjugate elementary divisors* $(\lambda - a \pm ib)^{e_c}$.

Hence by Theorem 1 we have

THEOREM 4. *Necessary and sufficient conditions, that two canonical matrices A_1 and A_2 be similar under a canonical transformation, are that*

(α) *the elementary divisors of the pencil $A_1 - \lambda E$ be the same as those of the pencil $A_2 - \lambda E$, and that*

(β) *the indices of all elementary divisors $(\lambda \pm 1)^{2k}$ and of all pairs of conjugate elementary divisors $(\lambda - a \pm ib)^k$, $a^2 + b^2 = 1$, be the same for both pencils.*

---

[10] Cf. John Williamson, "The equivalence of non-singular pencils of hermitian matrices in an arbitrary field," *American Journal of Mathematics*, vol. 57 (1935), pp. 484-485.

6. Normal form of a canonical matrix. In order to determine the normal form, to which a canonical matrix A may be reduced by a canonical transformation, it is only necessary, on account of Theorem 1, to reduce the associated skew symmetric matrix S of section 4 to the form G by a matrix R, and then to determine RQR^{-1}. As a first step we reduce each S_j of type b to the form G_j.

Type b_1. Since $e_j = 2m$, we may write

$$P_{e_j} = P_{2m} = \begin{pmatrix} P_m & L_m \\ 0 & P_m \end{pmatrix},$$

where P_m is of order m and is defined by (12), while all elements of L_m are zero except the element in the last row and first column which has the value p. By result III

$$S_j \approx \epsilon X = \epsilon \begin{pmatrix} 0 & C \\ -C' & 0 \end{pmatrix},$$

where C is a non-singular matrix of order m. Since

$$P_{2m} \epsilon X P'_{2m} = \epsilon X,$$

it is easily verified that

(56) $$P_m C' P'_m = C'.$$

Let $R = \begin{pmatrix} E & 0 \\ 0 & \epsilon (C')^{-1} \end{pmatrix}$. Then

$$R \epsilon X R' = \begin{pmatrix} 0 & E_m \\ -E_m & 0 \end{pmatrix} = G_m,$$

and

$$R P_{2m} R^{-1} = \begin{pmatrix} P_m & \epsilon L_m C' \\ 0 & (C')^{-1} P_m C' \end{pmatrix} = \begin{pmatrix} P_m & \epsilon M_m \\ 0 & (P'_m)^{-1} \end{pmatrix} \text{ by (54)},$$

where

(55) $$M_m = \begin{bmatrix} 0 & 0 & \cdot & 0 \\ 0 & 0 & \cdot & 0 \\ \cdot & & \cdot\ \cdot & 0 \\ p & -p & \cdot & (-1)^{m-1} p \end{bmatrix}.$$

Accordingly we have

Result III_a. In type b_1, when $e_j = 2m$, P_{e_j} may be replaced by

$$Z_{2m} = \begin{pmatrix} P_m & \epsilon M_m \\ 0 & (P'_m)^{-1} \end{pmatrix},$$

where M_m is defined by (55) *and $\epsilon = \pm 1$. Then $S_j \approx G_{2m}$.*

For example, if $m = 3$, and $\epsilon = 1$, P_{2m} may be replaced by

$$\begin{pmatrix} p & p & 0 & 0 & 0 & 0 \\ 0 & p & p & 0 & 0 & 0 \\ 0 & 0 & p & p & -p & p \\ 0 & 0 & 0 & p & 0 & 0 \\ 0 & 0 & 0 & -p & p & 0 \\ 0 & 0 & 0 & p & -p & p \end{pmatrix}.$$

If $p=1$, this last matrix is a canonical matrix of order six with the single elementary divisor $(\lambda-1)^6$.

Type b_2. Since $e_j=2m+1$ we may write

$$P_{e_j}=\begin{pmatrix} P_m & L_m \\ 0 & P_{m+1} \end{pmatrix},$$

where the only non-zero element of L_m is an element p in the last row and first column. By result IV

$$S_j \approx \epsilon Y,$$

and

$$Y=\begin{pmatrix} 0 & K \\ D & 0 \end{pmatrix},$$

where D is a non-singular $(m+1)$-rowed matrix, while K consists of the first m rows of $-D'$. As in the previous case we deduce that

$$P_{m+1}DP'_{m+1}=D. \tag{56}$$

Let $R=\begin{pmatrix} E_m & 0 \\ 0 & -\epsilon D^{-1} \end{pmatrix}$. Then

$$\begin{aligned} R\epsilon YR' &= \begin{pmatrix} E_m & 0 \\ 0 & -\epsilon D^{-1} \end{pmatrix}\begin{pmatrix} 0 & \epsilon K \\ \epsilon D & 0 \end{pmatrix}\begin{pmatrix} E_m & 0 \\ 0 & -\epsilon (D')^{-1} \end{pmatrix} \\ &= \begin{pmatrix} 0 & \epsilon K \\ -E_{m+1} & 0 \end{pmatrix}\begin{pmatrix} E_m & 0 \\ 0 & -\epsilon (D')^{-1} \end{pmatrix}. \end{aligned} \tag{57}$$

Since K is formed of the first m rows of $-D'$, $-K(D')^{-1}=(E_m,0)$. Further the first m elements of the last row of the product on the right of (57) are zero, while the last $m+1$ are the elements in the first row of $\epsilon(D^{-1})'$. The only element in the last column of D^{-1} different from zero is the last, which has the value $(-1)^{m-1}(i)$. Therefore the only element different from zero in the first row of $(D^{-1})'$ is the last, which has the value $(-1)^{m-1}(i)'=(-1)^m i$. Accordingly it follows from (57) that

$$R\epsilon YR'=\begin{pmatrix} 0 & E_m & 0 \\ -E_m & 0 & 0 \\ 0 & 0 & (-1)^m\epsilon i \end{pmatrix}. \tag{58}$$

But by (56)

$$(59)\qquad RP_{2m}R' = \begin{pmatrix} P_m & \epsilon N_m \\ 0 & (P'_{m+1})^{-1} \end{pmatrix} = F,$$

where the last row of N_m is p times the first row of $-D$, that is

$$(60)\qquad N_m = \begin{bmatrix} 0 & 0 & 0 & \cdot & 0 & 0 \\ 0 & 0 & 0 & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -\frac{ip}{2} & \frac{ip}{2} & \cdot & \cdot & (-1)^{m-2}\frac{ip}{2} & (-1)^{m-1}ip \end{bmatrix}.$$

It is not possible to proceed any further with the reduction without breaking up some of the two-rowed matrices into their component elements. Accordingly we write $N_m = K_m\alpha_1\alpha_2$, where α_1 and α_2 are matrices of a single column, and $(P'_{m+1})^{-1}$ in the form

$$(P'_{m+1})^{-1} = \begin{bmatrix} (P'_m)^{-1} & \gamma_1 & \gamma_2 \\ \delta_1 & a & b \\ \delta_2 & -b & a \end{bmatrix},$$

where γ_1, γ_2 are matrices of a single column and δ_1, δ_2 matrices of a single row. Then the matrix F in (59) becomes

$$\begin{bmatrix} P_m & \epsilon K_m & \epsilon\alpha_1 & \epsilon\alpha_2 \\ 0 & (P'_m)^{-1} & \gamma_1 & \gamma_2 \\ 0 & \delta_1 & a & b \\ 0 & \delta_2 & -b & a \end{bmatrix}.$$

If $\epsilon = (-1)^m$, so that $\epsilon(-1)^m i = i$, a simple interchange of rows and columns reduces the matrix on the right of (58) to

$$\begin{bmatrix} 0 & 0 & E_m & 0 \\ 0 & 0 & 0 & 1 \\ -E_m & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix} = G_{e_j}$$

and F to

$$(61)\qquad Z_{e_j} = \begin{bmatrix} P_m & \epsilon\alpha_1 & \epsilon K_m & \epsilon\alpha_2 \\ 0 & a & \delta_1 & b \\ 0 & \gamma_1 & (P_m) & \gamma_2 \\ 0 & -b & \delta_2 & a \end{bmatrix}.$$

On the other hand, if $\epsilon = (-1)^{m-1}$, since $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$,

the matrix on the right of (58) may be reduced to G_{e_j} and F to a matrix obtained from (61) by interchanging the subscripts 1 and 2 and b with $-b$. We therefore have

Result IV$_a$. In type b_2, when $e_j = 2m+1$, P_{2m+1} may be replaced by one of the forms (61). Then $S_j \approx G_{e_j}$.

By the above processes we may reduce S to the form $[G_1, G_2, \cdots, G_k]$, where $G_i = \begin{pmatrix} 0 & E_i \\ -E_i & 0 \end{pmatrix}$ and Q to the form $[Z_1, Z_2, \cdots, Z_k]$, where Z_i is determined from one of the results I, II, III$_a$ and IV$_a$. Let

$$Z_j = \begin{pmatrix} Z_{j,11} & Z_{j,12} \\ Z_{j,21} & Z_{j,22} \end{pmatrix}, \qquad (j = 1, 2, \cdots, k),$$

where $Z_{j,rs}$ is a square matrix of the same order as E_j. Then by a simple interchange of rows and the same interchange of columns, the matrix $[G_1, G_2, \cdots, G_k]$, may be reduced to G and at the same time $Z_1, Z_2, \cdots, Z_k$ to

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}. \tag{62}$$

The matrices A_{rs} in (62) are defined by

$$A_{rs} = [Z_{1,rs}, Z_{2,rs}, \cdots, Z_{k,rs}], \qquad (r, s = 1, 2). \tag{63}$$

The matrices (62) are uniquely determined, apart from a rearrangement of rows and the same rearrangement of the columns, by the elementary divisors of $A - \lambda E$ and the indices of these elementary divisors. Therefore we have

THEOREM 5. *Any canonical matrix is similar under a canonical transformation to one and (essentially) only one of the matrices* (62).

THE JOHNS HOPKINS UNIVERSITY.

CRITERIA FOR CERTAIN HIGHER CONGRUENCES.*

By Leonard Carlitz.

1. Introduction. The congruences in question are of the form

$$\sum_{i=0}^{s}(-1)^i\begin{bmatrix}s\\s-i\end{bmatrix}^{p^{n(i-1)}}u^{p^{ni}}\equiv M \pmod P, \tag{1.1}$$

where M and P are polynomials in an indeterminate x with coefficients in the Galois field $GF(p^n)$ of order p^n, and P is irreducible. As for the coefficients in the left member of (1.1), if we put

$$[k]=x^{p^{nk}}-x \qquad (k=0,1,2,\cdots), \tag{1.2}$$

$$\begin{aligned} F_k&=[k][k-1]^{p^n}\cdots[1]^{p^{n(k-1)}}, & F_0&=1,\\ L_k&=[k][k-1]\cdots[1], & L_0&=1, \end{aligned} \tag{1.3}$$

then we define

$$\begin{bmatrix}s\\i\end{bmatrix}=\frac{F_s}{F_iL_{s-i}^{p^{ni}}}, \qquad \begin{bmatrix}s\\0\end{bmatrix}=\frac{F_s}{L_s}, \qquad \begin{bmatrix}s\\s\end{bmatrix}=1. \tag{1.4}$$

Thus the polynomial in u occurring in (1.1) closely resembles the polynomial [1]

$$\psi_s(u)=\sum_{i=0}^{s}(-1)^{s-i}\begin{bmatrix}s\\i\end{bmatrix}u^{p^{ni}}, \tag{1.5}$$

which has the characteristic property

$$\psi_s(u)=\prod_{\deg E<s}(u-E), \tag{1.6}$$

the product extending over all polynomials E (including 0) of degree $< s$. A closer connection will appear below.

If now we put

$$M\equiv A^{p^{n(s-1)}} \pmod P,$$

as may always be done, we shall derive the following criterion [2] for the congruence (1.1): Let

* Received December 11, 1936.

[1] See *Duke Mathematical Journal*, vol. 1 (1936), pp. 139-142; this paper will be cited as DJ. For the congruence $\psi_s(u)\equiv M$, see *Bulletin of the American Mathematical Society*, vol. 41 (1935), pp. 907-914.

[2] For the case $s=1$, see DJ, pp. 164-168.

$$(1.7)\qquad \begin{aligned} P &= x^k + c_1x^{k-1} + \cdots + c_k, && (c_j \text{ in } GF(p^n)),\\ P' &= kx^{k-1} + (k-1)c_1x^{k-2} + \cdots + c_{k-1}, \end{aligned}$$

so that P' is the (formal) derivative of P. Assume $k > s$: *then the congruence* (1.1) *is solvable if and only if the product AP' is congruent* (mod P) *to a polynomial of degree* $< k - s$. If this condition is satisfied the congruence has precisely p^{ns} solutions.

2. The polynomials $g_s(u)$ and $f_s(u)$. We denote by $g_s(u)$ the polynomial in the left member of (1.1). Since, by (1.2),

$$[s] = [s-i] + [i]^{p^{n(s-i)}},$$

it is evident from (1.3) and (1.4) that

$$\begin{bmatrix} s \\ s-i \end{bmatrix} = \frac{[s]F_{s-1}^{p^n}}{F_{s-i}L_i^{p^{n(s-i)}}} = \frac{F_{s-1}^{p^n}}{F_{s-i-1}^{p^n}L_i^{p^{n(s-i)}}} + \frac{F_{s-1}^{p^n}}{F_{s-i}L_{i-1}^{p^{n(s-i)}}},$$

so that

$$(2.1)\qquad \begin{bmatrix} s \\ s-i \end{bmatrix} = \begin{bmatrix} s-1 \\ s-i-1 \end{bmatrix}^{p^n} + F_{s-1}^{p^n-1}\begin{bmatrix} s-1 \\ s-i \end{bmatrix},$$

for $0 < s - i \leqq s$: by properly defining our symbols we may assert that (2.1) holds also for $s - i = 0, s$. Then, by substituting in the left member of (1.1), we have

$$\begin{aligned} g_s(u) &= \sum_{i=0}^{s} (-1)^i \left\{ \begin{bmatrix} s-1 \\ s-i-1 \end{bmatrix}^{p^n} + F_{s-1}^{p^n-1}\begin{bmatrix} s-1 \\ s-i \end{bmatrix} \right\}^{p^{n(i-1)}} u^{p^{ni}} \\ &= \sum_{i=0}^{s-1} (-1)^i \begin{bmatrix} s-1 \\ s-i-1 \end{bmatrix}^{p^{ni}} u^{p^{ni}} \\ &\qquad - \sum_{i=0}^{s-1} (-1)^i \begin{bmatrix} s-1 \\ s-i-1 \end{bmatrix}^{p^{ni}} F_{s-1}^{(p^n-1)p^{ni}} u^{p^{n(i+1)}}, \end{aligned}$$

from which it follows that [3]

$$(2.2)\qquad g_s(u^{p^n}) = g_{s-1}^{p^n}(u) - g_{s-1}^{p^n}(F_{s-1}^{1-p^{-n}}u^{p^n}) = g_{s-1}^{p^n}(u - F_{s-1}^{1-p^{-n}}u^{p^n}).$$

If we define $g_0(u) = u$, it is clear that (2.2) holds for all $s \geqq 1$.

We next define the polynomial $f_s(u)$ by means of

$$(2.3)\qquad f_s(u) = \sum_{i=0}^{k-s} \frac{F_{i+s-1}}{F_{s-1}F_i^{p^{n(s-1)}}} u^{p^{ni}} \qquad (1 \leqq s \leqq k),$$

where k is some fixed integer > 0. Then we have

---

[3] The quantity $F_{s-1}^{p^{-n}}$ may be defined in terms of the symbol $x^{p^{-n}}$. Otherwise the formula (2.2) may be interpreted as a congruence (mod P), in which case no new symbol is required; this interpretation is sufficient for the application.

$$f_s(u) - F_{s-1}^{p^n-1} f_s^{p^n}(u) = \sum_{i=0}^{k-s} \frac{F_{i+s-1}}{F_{s-1}F_i^{p^{n(s-1)}}} u^{p^{ni}} - \sum_{i=0}^{k-s} \frac{F_{i+s-1}^{p^n}}{F_{s-1}F_i^{p^{ns}}} u^{p^{n(i+1)}}$$
$$= \sum \left\{ \frac{F_{i+s-1}}{F_{s-1}F_i^{p^{n(s-1)}}} - \frac{F_{i+s-2}^{p^n}}{F_{s-1}F_{i-1}^{p^{ns}}} \right\} u^{p^{ni}}$$
$$= \sum \frac{F_{i+s-2}^{p^n}}{F_{s-1}F_i^{p^{n(s-1)}}} \{[i+s-1] - [i]^{p^{n(s-1)}}\} u^{p^{ni}}$$
$$= \sum \frac{F_{i+s-2}^{p^n}}{F_{s-2}^{p^n} F_i^{p^{n(s-1)}}} u^{p^{ni}},$$

from which follows the formula

$$f_s(u^{p^n}) - F_{s-1}^{p^n-1} f_s^{p^n}(u^{p^n}) = f_{s-1}^{p^n}(u), \tag{2.4}$$

for $s \geqq 2$. Now by (2.3),

$$f_1(u^{p^n}) - F_0^{p^n-1} f_1^{p^n}(u^{p^n}) = \sum_{i=0}^{k-1} u^{p^{n(i+1)}} - \sum_{i=0}^{k-1} u^{p^{n(i+2)}}$$
$$= u^{p^n} - u^{p^{n(k+1)}} = (u - u^{p^{nk}})^{p^n}.$$

Thus if we define $f_0(u)$ by means of

$$f_0(u) = u - u^{p^{nk}}, \tag{2.5}$$

it is evident that (2.4) holds for all $s \geqq 1$.

Making use of the formulas (2.2) and (2.4), we now prove the identical congruence

$$f_s(g_s(u)) \equiv u - u^{p^{nk}} \pmod{P}, \tag{2.6}$$

where P is irreducible of degree k.

For $s = 0$, (2.6) follows at once from (2.5) and $g_0(u) = u$. For $s = 1$,

$$g_1(u) = u - u^{p^n}, \quad f_1(u) = u + u^{p^n} + \cdots + u^{p^{n(k-1)}}, \quad f_1(g_1(u)) = u - u^{p^{nk}},$$

so that (2.6) holds in this case also. We now assume that (2.6) holds up to and including the value $s - 1$. Clearly [4]

$$\{g_s(u^{p^n})\}^{1/p^n} = g^{p^{-n}}(u^{p^n})$$

is uniquely determined (mod P). Then, by (2.2) and (2.4), we have for $s \geqq 1$,

$$f_s(g_s(u^{p^n})) - F_{s-1}^{p^n-1} f_s^{p^n}(g_s(u^{p^n})) = f_{s-1}^{p^n}(g_s^{p^{-n}}(u^{p^n})) \tag{2.7}$$
$$= f_{s-1}^{p^n}\{g_{s-1}(u - F_{s-1}^{1-p^{-n}} u^{p^n})\}$$

$$\equiv \{(u - F_{s-1}^{1-p^{-n}} u^{p^n}) - (u - F_{s-1}^{1-p^{-n}} u^{p^n})^{p^{nk}}\}^{p^n}, \tag{2.8}$$

---

[4] Compare footnote 3.

since (2.6) is assumed to hold for $s-1$. We rewrite (2.8) in the form

$$u^{p^n} - u^{p^{n(k+1)}} - (F_{s-1}^{p^n-1} u^{p^{2n}} - F_{s-1}^{(p^n-1)p^{nk}} u^{p^{n(k+2)}})$$

or

$$u^{p^n} - u^{p^{n(k+1)}} - F_{s-1}^{p^n-1}(u^{p^n} - u^{p^{n(k+1)}})^{p^n}.$$

If now we compare this with the left member of (2.7) and replace u^{p^n} by u, we get

$$f_s(g_s(u)) \equiv cF_{s-1} + u - u^{p^{nk}} \pmod P,$$

where c is in $GF(p^n)$; but from the form of $f_s(u)$ and $g_s(u)$ it is clear that $c=0$, so that (2.5) holds for the value s.

In a similar way we may prove the identical congruence

$$g_s(f_s(u)) \equiv u - u^{p^{nk}} \pmod P, \tag{2.9}$$

which obviously holds for $s=0,1$. Indeed to prove (2.9) note that

$$\begin{aligned} g_s(f_s(u^{p^n})) &\equiv g_{s-1}^{p^n}\{f_s^{p^{-n}}(u^{p^n}) - F_{s-1}^{1-p^{-n}} f_s(u^{p^n})\} \\ &\equiv g_{s-1}^{p^n}\{f_{s-1}(u)\} \equiv (u - u^{p^{nk}})^{p^n} \pmod P, \end{aligned}$$

which completes the induction. From (2.6) and (2.9) follows

THEOREM 1. *For P irreducible of degree k, and $0 \leqq s \leqq k$,*

$$u - u^{p^{nk}} \equiv f_s(g_s(u)) \equiv g_s(f_s(u)) \pmod P.$$

It is of some interest to observe that (2.6) and (2.9) are equivalent, that is, either implies the other (without the use of formulas (2.2) and (2.4)). This is a consequence of the following

LEMMA. *If $f(u) = \Sigma\, \alpha_j u^{p^{nj}}$, $g(u) = \Sigma\, \beta_j u^{p^{nj}}$, and*

$$f(g(u)) \equiv u - u^{p^{nk}} \pmod P,$$

where P is irreducible of degree k, then also

$$g(f(u)) \equiv u - u^{p^{nk}} \pmod P.$$

Since the proof is very much like the proof of a similar theorem [5] proved elsewhere, it will be omitted here.

3. Factorization of $f_s(u)$. From the identity (1.6) it follows readily that

$$\pi_{k-s}(u) = (-1)^{k-s} \frac{L_{k-s}}{F_{k-s}} \psi_{k-s}(u) = u \prod_{\deg E < k-s} (1 - u/E), \tag{3.1}$$

[5] DJ, p. 152.

the product extending over all E (except 0) of degree $< k-s$. On the other hand, by (1.5) and the first of (3.1),

$$\pi_{k-s}^{p^{n(s-1)}}(uL_{k-1}) = L_{k-s}^{p^{n(s-1)}} \sum_{j=0}^{k-1} (-1)^j \frac{L_{k-1}^{p^{n(j+s-1)}}}{F_j^{p^{n(s-1)}} L_{k-s-j}^{p^{n(j+s-1)}}} u^{p^{n(j+s-1)}}.$$

Now, from (1.3), it is easily seen that

$$\left(\frac{L_{k-1}}{L_{k-s-j}}\right)^{p^{n(j+s-1)}} \equiv (-1)^{j+s-1} F_{j+s-1} \pmod P,$$

$$L_{k-s}^{p^{n(s-1)}} \equiv (-1)^{s-1} \frac{L_{k-1}^{p^{n(s-1)}}}{F_{s-1}},$$

so that

$$\pi_{k-s}^{p^{n(s-1)}}(uL_{k-1}) \equiv L_{k-1}^{p^{n(s-1)}} \sum_{j=0}^{k-s} \frac{F_{j+s-1}}{F_{s-1} F_j^{p^{n(s-1)}}} u^{p^{n(j+s-1)}},$$

and therefore,

$$\left\{\frac{1}{L_{k-1}} \pi_{k-s}(uL_{k-1})\right\}^{p^{n(s-1)}} \equiv f_s(u^{p^{n(s-1)}}) \pmod P. \tag{3.2}$$

Comparison of (3.2) and (3.1) leads to the factorization

$$f_s(u) \equiv u \prod_{\deg E < k-s} \left\{1 - \frac{uL_{k-1}^{p^{n(s-1)}}}{E^{p^{n(s-1)}}}\right\} \pmod P. \tag{3.3}$$

But [6] $L_{k-1} \equiv (-1)^{k-1} P'$, where P' is defined by (1.7); therefore we may put (3.3) in the form

$$f_s(u) \equiv u \prod_{\deg E < k-s} \left\{1 - \frac{uP'^{p^{n(s-1)}}}{E^{p^{n(s-1)}}}\right\} \pmod P. \tag{3.4}$$

THEOREM 2. *The polynomial $f_s(u)$ factors completely* $(\bmod P)$; *the roots are $(E/P')^{p^{n(s-1)}}$, where E ranges over the $p^{n(k-s)}$ polynomials of degree $< k-s$, and P' is the derivative of P.*

4. Criteria for solvability. If the congruence

$$g_s(u) \equiv M \pmod P \tag{4.1}$$

is assumed solvable, it follows from (2.6) that

$$f_s(M) \equiv f_s(g_s(u)) \equiv u - u^{p^{nk}} \pmod P, \tag{4.2}$$

for P irreducible of degree k. But by Fermat's Theorem, if A is any quantity $(\bmod P)$, $A^{p^{nk}} \equiv A$, so that (4.2) implies

$$f_s(M) \equiv 0 \pmod P; \tag{4.3}$$

[6] DJ, p. 166.

that is, (4.3) is a *necessary* condition that (4.1) be solvable. To show that this condition is also sufficient, we make use of Theorem 2 and formula (4.2). By Theorem 2 we have the factorization

$$f_s(g_s(u)) \equiv C \prod_{\delta} (g_s(u) - \delta) \pmod{P}, \tag{4.4}$$

where δ ranges over the roots of $f_s(\delta) \equiv 0$, and C is independent of u. If, now, we compare (4.4) with (4.2) and recall that $u^{p^{nk}} - u$ factors completely into linear factors, it is clear that for all δ (satisfying the congruence $f_s(\delta) \equiv 0$) the congruence $g_s(u) \equiv \delta$ is solvable; further, since $g_s(u) - \delta$ divides $u^{p^{nk}} - u$ it follows that the congruence in question has the maximum number of solutions. We may now state

THEOREM 3. *The congruence $g_s(u) \equiv M \pmod{P}$, where P is irreducible of degree $k > s$, is solvable if and only if $f_s(M) \equiv 0 \pmod{P}$. If this condition is satisfied, the congruence has precisely p^{ns} solutions.*

Now by Theorem 2, the roots of $f_s(u) \equiv 0$ are the quantities $(E/P')^{p^{n(s-1)}}$, where E is of degree $< k - s$. Thus it is necessary that M be congruent to one of these quantities. If then we replace M by $A^{p^{n(s-1)}}$ (clearly M *uniquely* determines A), we have the

THEOREM 4. *If P is irreducible of degree $k > s$, the congruence*

$$g_s(u) \equiv A^{p^{n(s-1)}} \pmod{P} \tag{4.5}$$

is solvable if and only if the product AP' is congruent $(\operatorname{mod} P)$ *to a polynomial of degree $< k - s$. If u_0 is a particular solution of* (4.5), *then the general solution is $u_0 + \rho$, where ρ ranges over the p^{ns} roots of $g_s(u) \equiv 0 \pmod{P}$.*

The second part of the theorem follows at once from the observation that $g_s(u) \equiv A^{p^{n(s-1)}}$ and $g_s(v) \equiv A^{p^{n(s-1)}}$ imply $g_s(u - v) \equiv 0$. We shall now determine the roots of $g_s(u) \equiv 0$.

5. The roots of $g_s(u) \equiv 0$. For $s = 1$, $g_1(u) = u - u^{p^n}$, and the roots are evidently the p^n elements of the $GF(p^n)$.

For $s = 2$, we make use of the recurrence (2.2). Thus we have the condition

$$g_2(u^{p^n}) \equiv g_1^{p^n}(u - F_1^{1-p^{-n}} u^{p^n}) \equiv 0. \tag{5.1}$$

Therefore, by the preceding paragraph,

$$u - F_1^{1-p^{-n}} u^{p^n} \equiv c \qquad (c \text{ in } GF(p^n)),$$

so that

$$F_1u^{p^n} - (F_1u^{p^n})^{p^n} \equiv c(x^{p^n} - x),$$

and, therefore, we have at once $F_1u^{p^n} \equiv cx + c'$, where c and c' are arbitrary elements of $GF(p^n)$. Thus by (5.1) the roots of $g_2(u) \equiv 0 \pmod P$ are furnished by $(cx + c')/F_1$.

For the case $s = 3$, we again employ (2.2)

$$g_3(u^{p^n}) \equiv g_2^{p^n}(u - F_2^{1-p^{-n}}u^{p^n}) \equiv 0; \tag{5.2}$$

then, as above, we get

$$u - F_2^{1-p^{-n}}u^{p^n} \equiv \frac{cx + c'}{F},$$

$$F_2u^{p^n} - (F_2u^{p^n})^{p^n} \equiv (cx^{p^n} + c')(x^{p^{2n}} - x),$$

from which follows easily

$$F_2u^{p^n} \equiv cx^{p^n+1} + c'(x^{p^n} + x) + c'',$$

where c, c', c'' are in $GF(p^n)$; thus by (5.2) the roots of $g_3(u) \equiv 0 \pmod P$ are furnished by

$$\frac{cx^{p^n+1} + c'(x^{p^n} + x) + c''}{F_2},$$

where c, c', c'' independently range over the elements of $GF(p^n)$.

It is now not difficult to determine the roots of $g_s(u) \equiv 0$. Let

$$\sigma_j = \sigma_j^{(s)} = \sigma_j(x, x^{p^n}, \cdots, x^{p^{n(s-1)}}) \tag{5.3}$$

denote the j-th elementary symmetric function of the quantities $x, x^{p^n}, \cdots, x^{p^{n(s-1)}}$; thus we have the identity

$$(t + x)(t + x^{p^n}) \cdots (t + x^{p^{n(s-1)}}) = \sum_{j=0}^{s} \sigma_j^{(s)} t^{s-j}. \tag{5.4}$$

We shall prove

THEOREM 5. *The p^{ns} roots of $g_s(u) \equiv 0 \pmod P$ are furnished by*

$$\rho = \frac{c_1\sigma_{s-1}^{(s-1)} + c_2\sigma_{s-2}^{(s-1)} + \cdots + c_s\sigma_0^{(s-1)}}{F_{s-1}}, \tag{5.5}$$

where the c_j independently range over the elements of $GF(p^n)$, and $\sigma_j^{(s-1)}$ is defined by (5.3).

The theorem is evidently true for $s = 1, 2, 3$. Assuming it to hold up to and including the value $s - 1$, we use (2.2)

$$g_s(u^{p^n}) \equiv g_{s-1}^{p^n}(u - F_{s-1}^{1-p^{-n}} u^{p^n}) \equiv 0.$$

Thus since the theorem is assumed true for the case $s-1$, we have at once

$$F_{s-1}u^{p^n} - (F_{s-1}u^{p^n}) = (c_1\sigma_{s-2}^{(s-2)} + \cdots + c_{s-1}\sigma_0^{(s-2)})^{p^n}(x^{p^{n(s-1)}} - x).$$

It is clear that, to complete the induction, it is only necessary to show that

(5.6) $$(c_1\sigma_{s-1}^{(s-1)} + \cdots + c_s\sigma_0^{(s-1)})^{p^n} - (c_1\sigma_{(s-1)}^{s-1} + \cdots + c_s\sigma_0^{(s-1)}) = (c_1\sigma_{s-2}^{(s-2)} + \cdots + c_{s-1}\sigma_0^{(s-2)})(x^{p^{n(s-1)}} - x).$$

From (5.4), it follows that

$$(t + x^{p^n})(t + x^{p^{2n}}) \cdots (t + x^{p^{ns}}) = \sum_{j=0}^{s} (\sigma_j^{(s)})^{p^n} t^{s-j};$$

combining this with (5.4) we have

$$(x^{p^{ns}} - x)\cdot(t + x^{p^n}) \cdots (t + x^{p^{n(s-1)}}) = \sum_{j=0}^{s} \{(\sigma_j^{(s)})^{p^n} - \sigma_j^{(s)}\} t^{s-j}.$$

But this implies

$$(\sigma_j^{(s)})^{p^n} - \sigma_j^{(s)} = (\sigma_{j-1}^{(s-1)})^{p^n}(x^{p^{ns}} - x);$$

in this formula replace s by $s-1$ and (5.6) follows immediately. This completes the proof of the theorem.

As an immediate corollary of Theorem 5 and the latter part of Theorem 4, we state

THEOREM 6. *If u_0 is a particular solution of $g_s(u) \equiv A^{p^{n(s-1)}} \bmod P)$, then the general solution is $u_0 + \rho$, where ρ is determined by* (5.5).

6. Some extensions. If $f_s(A^{p^{n(s-1)}}) \equiv 0 \pmod P$, it is clear from (2.4) that also $f_{s-1}(A^{p^{n(s-2)}}) \equiv 0$; however, the converse is not true in general. Thus it may happen that the congruence

$$g_s(u) \equiv A^{p^{n(s-1)}} \pmod P$$

is not solvable, while the congruence

$$g_{s-1}(u) \equiv A^{p^{n(s-2)}} \pmod P$$

is solvable. In this case it is easily seen that $g_s(u) - A^{p^{n(s-1)}}$ breaks up $(\bmod P)$ into a product of $p^{n(s-1)}$ factors each of degree p^n. This follows from the formula

(6.1) $$g_s(u) - A^{p^{n(s-1)}} = \{g_{s-1}(u^{p^{-n}} - F_{s-1}^{1-p^{-n}}u) - A^{p^{n(s-2)}}\}^{p^n},$$

which is an immediate consequence of (2.2). Assume $f_{s-1}(A^{p^{n(s-2)}}) \equiv 0$, so that by Theorem 3 we have the factorization

$$g_{s-1}(u) - A^{p^{n(s-2)}} \equiv (-1)^{s-1} \frac{F_{s-1}}{L_{s-1}} \prod_{\delta} (u - \delta) \pmod P,$$

where δ ranges over the $p^{n(s-1)}$ roots of $g_{s-1}(u) \equiv A^{p^{n(s-2)}}$. Substitution in (6 1) gives

$$(6\ 2)\quad g_s(u) - A^{p^{n(s-1)}} \equiv (-1)^s \left(\frac{F_{s-1}}{L_{s-1}}\right)^{p^n} \prod_{\delta} (F_{s-1}^{p^n-1} u^{p^n} - u + \delta^{p^n}) \pmod P.$$

More generally if we assume only $f_r(A^{p^{n(r-1)}}) \equiv 0 \pmod P$, where $r < s$, then we may show that $g_s(u) - A^{p^{n(s-1)}}$ factors $(\mathrm{mod}\, P)$ into a product of p^{nr} polynomials, each of degree $p^{n(s-r)}$. Thus formula (6.2) is the special case $r = s - 1$. We now show that in the general case we have a factorization of the form

$$(6\ 3)\qquad g_{r+s}(u) - A^{p^{n(r+s-1)}} \equiv (-1)^{r+s} \left(\frac{F_r}{L_r}\right)^{p^{ns}} \prod_{\delta} \{G_{r,s}(u) - (-1)^s \beta^{p^{ns}}\} \pmod P,$$

where $G_{r,s}(u)$ is a linear[7] polynomial of degree p^{ns}, and the product extends over the p^{nr} roots of $g_r(\beta) \equiv A^{p^{n(r-1)}}$.

Let r be fixed; the formula (6.3) is obviously true for $s = 0$. According to (6.2), the formula holds for $s = 1$. Assume that (6.3) holds up to and including the value $s - 1$. Then by (6.1) we have

$$\begin{aligned} g_{r+s}(u) - A^{p^{n(r+s-1)}} &\equiv \{g_{r+s-1}(u^{p^{-n}} - F_{r+s-1}^{1-p^{-n}} u) - A^{p^{n(r+s-2)}}\}^{p^n} \\ &\equiv (-1)^{r+s-1} \left(\frac{F_r}{L_r}\right)^{p^{ns}} \prod_{\beta} \{G_{r,s-1}^{p^n}(u^{p^{-n}} - F_{r+s-1}^{1-p^{-n}} u) - (-1)^{s-1}\beta^{p^{ns}}\} \\ &\equiv (-1)^{r+s} \left(\frac{F_r}{L_r}\right)^{p^{ns}} \prod_{\beta} \{G_{r,s}(u) - (-1)^s \beta^{p^{ns}}\}, \end{aligned}$$

thus completing the induction. It is also evident from the above that the polynomial $G_{r,s}(u)$ satisfies the recurrence

$$(6\ 4)\qquad G_{r,s}(u) = G_{r,s-1}^{p^n}(F_{r+s-1}^{1-p^{-n}} u - u^{p^{-n}}), \qquad G_{r,0}(u) = u.$$

We have therefore proved

THEOREM 7. *If $f_r(A^{p^{n(r-1)}}) \equiv 0 \pmod P$, the polynomial $g_{r+s}(u) - A^{p^{n(r+s-1)}}$ has the factorization (6.3), where $G_{r,s}(u)$ is determined by (6.4).*

---

[7] That is, of the form $\Sigma a_j u^{p^{nj}}$.

We shall now show that the polynomial $G_{r,s}(u) - (-1)^s \beta^{p^{ns}}$ occurring in the right member of (6.3) can in general be factored further (mod P); irreducibility occurs only in the case $n = 1 = s$, $f_{r+1}(A^{p^{nr}}) \not\equiv 0$ (mod P).

It is convenient to deal with the left member of (6.3). Let

$$(6.5) \qquad h(u) = g_s(u) - M,$$

where M is arbitrary (mod P). Then by (2.6),

$$(6.6) \qquad f_s\{h(u)\} = f_s\{g_s(u) - M\} \equiv u - u^{p^{nk}} - f_s(M) \pmod P.$$

In the next place,

$$(6.7) \qquad \begin{aligned} f_s\{h(u)\} - f_s^{p^{nk}}\{h(u)\} &\equiv u - 2u^{p^{nk}} + u^{p^{2nk}} - \{f_s(M) - f_s^{p^{nk}}(M)\} \\ &\equiv u - 2u^{p^{nk}} + u^{p^{2nk}} \pmod P, \end{aligned}$$

since for arbitrary M, $M^{p^{nk}} \equiv M$ (mod P). Now put

$$\begin{aligned} u_1 &= u - u^{p^{nk}}, \\ u_2 &= u_1 - u_1^{p^{nk}} = u - 2u^{p^{nk}} + u^{p^{2nk}}, \\ &\cdots\cdots\cdots\cdots \\ u_p &= u_{p-1} - u_{p-1}^{p^{nk}} = u - u^{p^{nkp}}. \end{aligned}$$

Clearly u_{j+1} is a multiple of u_j ($j = 1, \cdots, p-1$). Thus it follows that u_2 divides u_p. Therefore by (6.6) and (6.7), the polynomial $h(u)$ is a divisor of u_p:

$$(6.8) \qquad h(u) \mid u - u^{p^{nkp}}.$$

Now on the other hand since the set of residues (mod P) form a finite field $GF(p^{nk})$, it follows from a well known theorem that all the irreducible divisors (mod P) of u_p are of degree 1 or p. Therefore by (6.8) the same is true of $h(u)$. On the other hand it is evident from Theorem 6 that if $g_s(u) - M$ has one linear factor, then it has p^{ns} linear factors. This proves the following

THEOREM 8. *For arbitrary M, the polynomial $g_s(u) - M$* (mod P) *either factors completely into linear factors, or else is a product of $p^{n(s-1)}$ irreducible polynomials each of degree p.*

By Theorem 5, if $f_r(M) \not\equiv 0$ (mod P), the polynomial $g_s(u) - M$ has no linear factor. Hence by the preceding theorem we have

THEOREM 9. *If $f_s(M) \not\equiv 0$* (mod P), *then the polynomial $g_s(u) - M$ is a product* (mod P) *of $p^{n(s-1)}$ irreducible factors each of degree p.*

Suppose now in (6.5) we put $M \equiv A^{p^{n(r+s-1)}}$. Then comparison with (6.3) leads at once to

THEOREM 10. *Let* $f_r(A^{p^{n(r-1)}}) \equiv 0$, $f_{r+1}(A^{p^{nr}}) \not\equiv 0 \pmod P$; *let* β *be a root of* $g_r(\beta) \equiv A^{p^{n(r-1)}}$. *Then the polynomial* $G_{r,s}(u) - (-1)^s\beta^{p^{ns}}$ *occurring in the right member of* (6.3) *is a product* $(\bmod P)$ *of* $p^{n(s-1)}$ *irreducible polynomials each of degree* p.

In particular for $n = 1$, $s = 1$, we get

THEOREM 11. *If the hypotheses of Theorem* 10 *hold, and in addition* $n = s = 1$, *then the polynomial*

$$G_{r,1}(u) + \beta^p = F_r^{p-1}u^p - u + \beta^p$$

is irreducible $(\bmod P)$.

It is not difficult to prove this theorem directly.

DUKE UNIVERSITY,
DURHAM, NORTH CAROLINA.

ON A TRIGONOMETRICAL SERIES OF RIEMANN.*

By Aurel Wintner.

In his paper on Riemann integrals and trigonometrical series, Riemann[1] considers the series

(1) $$\sum_{k=1}^{\infty} \frac{\psi(kx+\frac{1}{2})}{k}$$

and

(2) $$\sum_{n=1}^{\infty} \frac{c(n)}{n} \sin 2\pi n x,$$

where

(3) $$\psi(x) = \begin{cases} x-[x]-\frac{1}{2} & \text{if } x \neq [x] \\ 0 & \text{if } x=[x] \end{cases}$$

and

(4) $$c(n) = \sum_{d|n} (-1)^d,$$

the summation in (4) being extended over the $d(n)$ divisors d of n. Riemann's aim in considering the series (1), (2) is to illustrate the limitations to which his definition of an integral subjects the theory of Fourier series. In fact, Riemann observes that if x is rational, both series (1), (2) are convergent and represent the same value, while the function defined by these series on the set of rational numbers is a non-bounded function on the set of those rational numbers which are contained in any fixed interval.

This statement of Riemann has recently been verified by Chowla and Walfisz[2] who discussed the series (1) and (2) for irrational values x as well. They proved, among other things, that the trigonometrical series (2) is almost everywhere convergent and represents almost everywhere the sum of the series (1). That the series (1) is almost everywhere convergent, is an obvious consequence of Khintchine's metrical results concerning diophantine approximations.

The object of the present note is an approach to Riemann's series from the point of view of Lebesgue's theory. It will be seen that the *trigonometrical* series (2) is a *Fourier* series in the sense of Lebesgue and belongs to the

* Received April 28, 1937.

[1] B. Riemann, *Gesammelte mathematische Werke*, 2nd edition, Leipzig, 1892, p. 263.

[2] S. Chowla and A. Walfisz, "Ueber eine Riemannsche Identität," *Acta Arithmetica*, vol. 1 (1935), pp. 87-112.

function defined by (1), i. e., that the odd periodic function (1) is integrable in the sense of Lebesgue and has the coefficients of (2) as Fourier constants. It will also be shown that not only the function (1) but also every positive power of it is integrable in the sense of Lebesgue, i. e., that *the function* (1) *is of class* L^p *for arbitrarily large* p, although this function is non-bounded in every interval (also if one discards sets of measure zero). It has, perhaps, a historical interest that the Lebesgue theory of integration and of Fourier series applies without difficulty to the example by means of which Riemann himself illustrated the limitations of his theory.

That (2) *is almost everywhere convergent* to the function (1), will turn out to be an immediate consequence of the fact[3] that if a Fourier series

$$(5_1) \qquad f(x) \sim \Sigma\,(a_n \cos nx + b_n \sin nx)$$

is such that

$$(5_2) \qquad \Sigma\, n^\delta (a_n{}^2 + b_n{}^2) < +\infty, \text{ where } \delta > 0,$$

then (5_1) is convergent almost everywhere to the function $f(x)$. The exceptional x-set of measure zero is, according to Chowla and Walfisz,[2] such that its elements x essentially depend on the arithmetical structure of the number x.

First, it will be shown that the series

$$(6) \qquad \sum_{m=1}^{\infty} \frac{\psi(mx)}{m}$$

is convergent in the mean, i. e., that

$$(7) \qquad \lim_{\substack{n\to\infty \\ j\to\infty}} \int_0^1 \{f_{n+j}(x) - f_n(x)\}^2\, dx = 0,$$

where $f_n(x)$ denotes the odd periodic R-integrable function

$$(8) \qquad f_n(x) = \sum_{m=1}^{n} \frac{\psi(mx)}{m}.$$

Since the Fourier series of the Bernoulli polynomial (3) is

$$(9) \qquad \psi(x) \sim -\frac{1}{\pi} \sum_{k=1}^{\infty} \frac{1}{k} \sin 2\pi kx,$$

it is seen from (8) that

---

[3] According to Kolmogoroff and Seliverstoff, one can replace n^δ in (5_2) by $\log n$; cf. A. Kolmogoroff and G. Seliverstoff, "Sur la convergence des séries de Fourier," *Rendiconti della Reale Accademia Nazionale dei Lincei*, ser. 6, vol. 3 (1926), pp. 307-310.

(10) $$f_n(x) \sim -\frac{1}{\pi}\sum_{k=1}^{\infty}\left(\frac{1}{k}\sum_{d|k}^{1\leqq d\leqq n} 1\right)\sin 2\pi kx,$$

where the inner sum is extended over those divisors d of k which are not greater than n. Thus, for every positive integer j,

$$f_{n+j}(x) - f_n(x) \sim -\frac{1}{\pi}\sum_{k=1}^{\infty}\left(\frac{1}{k}\sum_{d|k}^{n<d\leqq n+j} 1\right)\sin 2\pi kx.$$

Hence it is seen from Parseval's relation that (7) is equivalent to

$$\lim_{\substack{n\to\infty \\ j\to\infty}} \sum_{k=1}^{\infty}\left(\frac{1}{k}\sum_{d|k}^{n<d\leqq n+j} 1\right)^2 = 0,$$

i. e., to

$$\lim_{n\to\infty} \sum_{k=n+1}^{\infty}\left(\frac{1}{k}\sum_{d|k} 1\right)^2 = 0.$$

Since $\sum_{d|k} 1$ is the number $d(k)$ of divisors of k, it follows that one merely has to show the convergence of the series

$$\sum_{k=1}^{\infty}\frac{d(k)^2}{k^2}.$$

Now

(11) $$\sum_{k=1}^{\infty}\frac{d(k)^2}{k^\beta}$$

is convergent for every $\beta > 1$, since $|\zeta(\sigma+it)|^4$ has the finite mean value

$$\lim_{T\to\infty}\frac{1}{T}\int_1^T |\zeta(\sigma+it)|^4\,dt = \sum_{k=1}^{\infty}\frac{d(k)^2}{k^{2\sigma}}$$

for [4] every $\sigma > \frac{1}{2}$.

By a well-known theorem of Fischer, the relation (7) just proved implies the existence of a function $f(x)$ which belongs to the class L^2 and is such that

(12) $$\lim_{n\to\infty}\int_0^1 \{f(x) - f_n(x)\}^2\,dx = 0.$$

Since the functions (8) are the partial sums of the series (6), the relation (12)

[4] Cf., e. g., E. C. Titchmarsh, *The zeta-function of Riemann*, Cambridge, 1930, pp. 38-41.

may be expressed by saying that the series (6) converges in the mean to a function $f(x)$ of class L^2.

It is well known that (12) implies the existence of a subsequence of $\{f_r(x)\}$ such that this subsequence of $\{f_n(x)\}$ tends almost everywhere to $f(x)$. In other words, there exists an increasing sequence $\{\mu_h\}$ of positive integers such that if one unites the first μ_1, then the next μ_2 terms of the series (6), and so on, the resulting "bracketed" series converges almost everywhere to $f(x)$. Actually, the introduction of the brackets is superfluous in view of Khintchine's result referred to above. This fact will not be needed in what follows. For (12) in itself allows one to consider (6) as the definition of an odd periodic function $f(x)$ of class L^2, this function being undetermined on a set of measure zero.

Since the functions (8) tend in the mean to the function (6) of class L^2, the k-th Fourier constant of (8) tends, as $n \to \infty$, to the k-th Fourier constant of (6) for every fixed k. Hence it is seen from (10) that the k-th Fourier constant of (6) is

$$\lim_{n\to\infty} -\frac{1}{\pi k} \sum_{d|k}^{1\leqq d\leqq n} 1 = -\frac{1}{\pi k} \sum_{d|k} 1 = \frac{d(k)}{-\pi k}.$$

In other words,

$$\sum_{m=1}^{\infty} \frac{\psi(mx)}{m} \sim -\frac{1}{\pi} \sum_{k=1}^{\infty} \frac{d(k)}{k} \sin 2\pi kx. \tag{13}$$

It follows that the function (6) is of class L^p not only for $p = 2$ but for arbitrarily large p. In fact, on applying to (13) Hausdorff's extension of the Fischer-Riesz theorem, it is seen that (6) is of class L^p for every p if

$$\sum_{k=1}^{\infty} \left(\frac{d(k)}{k}\right)^{p/(p-1)}$$

is convergent for every $p \geqq 2$. Thus it is sufficient to show that

$$\sum_{k=1}^{\infty} \left(\frac{d(k)}{k}\right)^{1+\epsilon}$$

is convergent for every positive $\epsilon \leqq 1$. Since $d(k) \geqq 1$, it follows that it is sufficient to know the convergence of

$$\sum_{k=1}^{\infty} \frac{d(k)^2}{k^{1+\epsilon}}$$

for every $\epsilon > 0$. Since (11) is convergent for every $\beta > 1$, the proof is complete.

The convergence of (11) for $\beta > 1$ also implies that

$$\sum_{k=1}^{\infty} k^{\delta}\left(\frac{d(k)}{k}\right)^{2}$$

is convergent for sufficiently small values of $\delta > 0$. Hence, on comparing (13) with (5_2), it is seen from the criterion (5_1) that the Fourier series (13) is convergent almost everywhere. Since the arithmetical means of a Fourier series tend almost everywhere to the function to which it belongs, the sum of the Fourier series (13) is almost everywhere equal to the function (6).

The above results concern not Riemann's series (1), (2) but the function (6). The transition to Riemann's series can be based on the Bernoulli identity

$$\psi(x+\tfrac{1}{2}) = \psi(2x) - \psi(x), \tag{14}$$

which is obvious from the definition (3).

First, since (6) is an odd periodic function of class L^p, where p is arbitrarily large, the same holds for the function

$$\sum_{m=1}^{\infty} \frac{\psi(m2x)}{m}.$$

It follows, therefore, from Hölder's inequality that the function

$$\sum_{m=1}^{\infty} \frac{\psi(2mx)}{m} - \sum_{m=1}^{\infty} \frac{\psi(mx)}{m} \tag{15}$$

is of class L^p for every p. Now (15) is, in view of (14), identical with Riemann's function (1). Furthermore, it is seen from (13) that the Fourier series of the difference (15) is

$$\frac{-1}{\pi}\sum_{k=1}^{\infty} \frac{d(k)}{k}\sin 4\pi kx + \frac{1}{\pi}\sum_{k=1}^{\infty} \frac{d(k)}{k}\sin 2\pi kx. \tag{16}$$

Now $d(k) = \sum_{d|k} 1$, so that the Fourier series (16) is, in view of (4), identical with Riemann's trigonometrical series (2). Accordingly, Riemann's function (1) is of class L^p for every p and has the Fourier series

$$\sum_{m=1}^{\infty} \frac{\psi(mx+\frac{1}{2})}{m} \sim -\frac{1}{\pi}\sum_{n=1}^{\infty}\left(\frac{1}{n}\sum_{d|n}(-1)^{d}\right)\sin 2\pi nx. \tag{17}$$

Finally, the Fourier series (17) converges almost everywhere to the function

(1). In fact, (1) is identical with (15), while the Fourier series (13) converges almost everywhere to the function (6).

While $(\Psi(x))^n$, where $\Psi(x)$ denotes Riemann's function (1), is for every n integrable in the sense of Lebesgue, the function $\Psi(x)$ lies in every sense outside of the range of Riemann's integration theory. In fact, if ι is any subinterval of the interval $0 \leqq x \leqq 1$, there cannot exist a constant $M = M_\iota$ such that $|\Psi(x)| \leqq M$ almost everywhere in ι. For suppose, if possible, that there exists an $M = M_\iota$ for some ι. Then, if λ is any closed interval contained in the interior of ι, one can find a constant $K = K_\lambda$ such that $|S_n(x)| \leqq K$ for every x in λ and for every n, where the $S_n(x)$ denote the arithmetical means of the Fourier series (17) of $\Psi(x)$. In particular, the $S_n(x)$ are uniformly bounded on the set of rational x contained in λ. This clearly contradicts the fact mentioned by Riemann [1] and verified by Chowla and Walfisz.[2] It follows, in particular, that if a function $F(x)$ is almost everywhere equal to Riemann's function (1), then $F(x)$ is discontinuous at every x.

THE JOHNS HOPKINS UNIVERSITY.

ON DIVERGENT INFINITE CONVOLUTIONS.*

By E. R. van Kampen and Aurel Wintner.

Introduction. It is known that the distribution theory of sums of independent random variables can be developed from several points of view which are, however, in the main, equivalent. One, and possibly the most general, approach is represented by Kolmogoroff's axiomatic treatment[1] of questions of probability distribution. This approach applies, in particular, to the problem of probable convergence of series of independent random variables, as first solved by Khintchine and Kolmogoroff.[2] It also implies the treatment based on the Lebesgue measure theory of infinite product spaces, as developed by Steinhaus, Littlewood, Paley and Zygmund, Jessen, and others.[3]

It is known[4] that the main result of Khintchine and Kolmogoroff, which is based on the notion of "equivalent series," can be formulated also in terms of infinite convolutions. The results of the present paper concern infinite convolutions and imply, among other things, certain facts which are equivalent to theorems concerning the divergence problem of series of independent random variables. In particular, the results imply essential refinements of certain facts indicated by Lévy.[5] Since Lévy's statements will not be used, the following considerations imply detailed proofs for them.

Theorem 1, which applies not only to convolution sequences, seems to have an independent interest. Theorems 3 and 5 delimit all possibilities which can

* Received March 31, 1937.

[1] A. Kolmogoroff, "Grundbegriffe der Wahrscheinlichkeitsrechnung," *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 2 (Berlin, 1933), no. 3.

[2] A. Khintchine and A. Kolmogoroff, "Ueber Konvergenz von Reihen, deren Glieder durch den Zufall bestimmt werden," *Recueil de la Societé Mathématique de Moscou*, vol. 32 (1925), pp. 668-677; A. Kolmogoroff, "Ueber die Summen durch den Zufall bestimmter zufälliger Grössen," *Mathematische Annalen*, vol. 99 (1928), pp. 309-319; vol. 102 (1930), pp. 484-488.

[3] Cf. P. Lévy, "Sur quelques points de la théorie des probabilités dénombrables," *Annales de l'Institut Henri Poincaré*, vol. 6 (1936), pp. 153-184, where further references are given.

[4] B. Jessen and A. Wintner, "Distribution functions and the Riemann zeta function," *Transactions of the American Mathematical Society*, vol. 38 (1935), pp. 48-88, more particularly 84-86.

[5] P. Lévy, "Sur les séries dont les termes sont des variables eventuelles indépendentes," *Studia Mathematica*, vol. 3 (1931), pp. 119-155, more particularly chap. I; cf. also the corrections on p. 337 of vol. 3 (1934) of the *Annali della R. Scuola Normale Superiore di Pisa.*

actually occur in case of a divergent infinite convolution. Theorem 9 describes what can happen to a divergent or not absolutely convergent infinite convolution upon a reordering of its "factors." Due to the correspondence alluded to above,[4] a part of Theorem 3 can be interpreted as a manifestation of the famous "0 or 1" principle.[6] The total content of Theorem 3, which concerns infinite convolutions, cannot conveniently be formulated in terms of the probable convergence or divergence of sums of independent random variables, or in terms of a Lebesgue measure of an infinite product space.

All distribution problems under consideration will be assumed to be one-dimensional, so that the random variables are real numbers.

Metric. In what follows, Greek letters $\phi, \rho, \cdots$ will denote monotone non-decreasing functions of a real variable x which remain bounded as $x \to \pm \infty$. The case of a constant function α, i. e., the case where $\phi(x) = \alpha$ for every x and for some real number α, is not excluded.

(I) For a given function $y = \phi(x)$, the symbol $]\phi[$ will denote the bounded open interval $\phi(-\infty) < y < \phi(+\infty)$ or the point $y = \phi(-\infty)$ according as $\phi(-\infty) < \phi(+\infty)$ or $\phi(-\infty) = \phi(+\infty)$, i. e., according as $\phi(x)$ is not or is a constant function.

It will be convenient to consider a function $y = \phi(x)$ as a Jordan curve in an (x, y)-plane. This is made possible by adjoining the point $(x, y) = (x, \phi(x))$ to the segment constituted by the set of points (x, y), where y describes the closed interval $\phi(x-0) \leqq y \leqq \phi(x+0)$, if x is a discontinuity point of ϕ. Thus two functions, ϕ and ψ, determine the same Jordan curve if and only if the two functions are equal at their continuity points, i. e., if $\phi(x+0) = \psi(x+0)$ and/or $\phi(x-0) = \psi(x-0)$ for every x. In this case the functions ϕ and ψ will be considered as identical. Correspondingly, a sequence of functions $\phi_n(x)$ is said to be convergent if there exists a function $\phi(x)$ such that $\phi_n(x) \to \phi(x)$ holds at every continuity point x of $\phi(x)$. The signs $=$ and $\to$ will only be used in the sense just defined. By a classical theorem of Helly, $\phi_n \to \phi$ whenever $\phi_n(x) \to \phi(x)$ holds for a dense set of values x.

For a given number $\epsilon > 0$, let the ϵ-strip about the Jordan curve $y = \phi(x)$ be defined as the set of those points of the (x, y)-plane whose distance from at least one point of the Jordan curve is less than ϵ.

(II) For two functions $y = \phi_1(x)$, $y = \phi_2(x)$, let $|\phi_1; \phi_2|$ or $|\phi_1(x); \phi_2(x)|$ denote the greatest lower bound of those $\epsilon > 0$ for which every

[6] A. Kolmogoroff, *loc. cit.* [1], pp. 60-61.

point of the Jordan curve $y = \phi_1(x)$ is contained in the ϵ-strip about the Jordan curve $y = \phi_2(x)$.

Thus $|\phi_1;\phi_2|$ is a non-negative number not less than

$$\operatorname{Max}\big(|\phi_1(+\infty)-\phi_2(+\infty)|,\ |\phi_1(-\infty)-\phi_2(-\infty)|\big)$$

and not greater than

$$\operatorname{Max}\big(\phi_1(+\infty)-\phi_1(-\infty),\ \phi_2(+\infty)-\phi_2(-\infty)\big).$$

It is easily verified that

$$(1_1)\qquad |\phi_1;\phi_2| = 0 \text{ if and only if } \phi_1 = \phi_2;$$

$$(1_2)\qquad |\phi_1;\phi_2| = |\phi_2;\phi_1|;$$

$$(1_3)\qquad |\phi_1;\phi_2| \leqq |\phi_1;\phi_3| + |\phi_3;\phi_2|.$$

This means that (II) defines a metrization of the space of all functions $\phi(x)$, if the sign $\phi_1 = \phi_2$ is defined as above.[7] This metrization is easily seen to be a complete metrization, in the sense that

$$(2)\qquad \lim_{\substack{n=\infty\\ m=\infty}} |\phi_n;\phi_m| = 0 \text{ if and only if } \lim_{n=\infty} |\phi_n;\phi| = 0 \text{ for a } \phi.$$

On the other hand, it is not true that convergence with reference to the metrization defined by (II) is equivalent to convergence represented above by the symbol $\phi_n \to \phi$. This is shown by the example $\phi_n(x) = \operatorname{sgn}(x+n)$, where $\phi_n \to \phi$ does, but $|\phi_n;\phi| \to 0$ does not, hold for $\phi(x) \equiv 1$. The general situation is easily seen to be this:

(III) The three conditions

$$(3_1)\ \phi_n \to \phi; \qquad (3_2)\ \phi_n(+\infty) \to \phi(+\infty); \qquad (3_3)\ \phi_n(-\infty) \to \phi(-\infty)$$

are necessary and sufficient for $|\phi_n;\phi| \to 0$.

The function set $\|\phi_n\|$. Two functions $\phi_1(x)$, $\phi_2(x)$ will be said to be congruent if there exists a number c such that the two functions $\phi_1(x)$, $\phi_2(x-c)$ are identical.

(IV) For a given sequence $\{\phi_n\}$ of functions $\phi_n(x)$, let $\|\phi_n\|$ denote the set of those functions $\rho(x)$ for which one can choose a sequence $\{c_n\}$ of numbers c_n such that

$$(4)\qquad \phi_n(x-c_n) \to \rho(x), \qquad n \to \infty,$$

holds at every continuity point x of ρ.

[7] For another metrization, cf. P. Lévy, *loc. cit.* [5], pp. 339-341.

It is clear from this definition that

(5) $\quad \| \phi_n \| \subset \| \psi_m \|$ whenever $\{\psi_m\}$ is a subsequence of $\{\phi_n\}$,

and that

(6) $\quad$ if $\rho(x) \subset \| \phi_n \|$, then $\rho(x - c) \subset \| \phi_n \|$ for every c.

(V) If $\phi_n{}^1$ and $\phi_n{}^2$ are congruent for $n = 1, 2, \cdots$, then the two function sets $\| \phi_n{}^1 \|$, $\| \phi_n{}^2 \|$ are identical.

This is clear from the definitions.

LEMMA 1. *If ρ^1 and ρ^2 are contained in a function set $\| \phi_n \|$, then either the two functions ρ^1, ρ^2 are congruent or the two point sets $]\rho^1[$, $]\rho^2[$, defined under* (I), *have no point in common.*

Proof. Suppose, if possible, that ρ^1, ρ^2 are not congruent and $]\rho^1[$, $]\rho^2[$ do have a point in common. Then, on interchanging, if necessary, ρ^1 and ρ^2, there clearly exists an $x = x_0$ for which

(7) $$\rho^1(-\infty) < \rho^2(x_0) < \rho^1(+\infty).$$

Since $\rho^1 \subset \| \phi_n \|$ and $\rho^2 \subset \| \phi_n \|$, there exist two sequences of numbers, say $\{c_n{}^1\}$ and $\{c_n{}^2\}$, such that

(8) $$\phi_n(x - c_n{}^1) \to \rho^1(x), \qquad \phi_n(x - c_n{}^2) \to \rho^2(x).$$

Now the sequence of the differences $c_n{}^1 - c_n{}^2$ contains a subsequence which tends either to a finite limit c or to $-\infty$ or to $+\infty$. In the first case (8) implies that $\rho^1(x) = \rho^2(x - c)$, which is a contradiction, since ρ^1 and ρ^2 are not congruent, by hypothesis. In the second case (8) clearly implies that $\rho^2(x) \leqq \rho^1(-\infty)$. Since this contradicts (7), and since the third case can be treated in the same way as the second case, the proof of Lemma 1 is complete.

LEMMA 2. *If $\{\rho_m\}$ is a sequence of functions contained in a function set $\| \phi_n \|$, then $\| \rho_m \| \subset \| \phi_n \|$.*

Proof. If $\sigma \subset \| \rho_m \|$ and σ is not a constant function, then Lemma 1 implies that $\sigma(x)$ and $\rho_m(x)$ are congruent for every sufficiently large m, so that $\sigma \subset \| \phi_n \|$, by (6). Hence it is sufficient to prove that every constant function α contained in $\| \rho_m \|$ is contained in $\| \phi_n \|$. In view of (V), one can assume without loss of generality that $\rho_m(x) \to \alpha$ as $m \to \infty$. Then there exists for every $\epsilon > 0$ and for every $t > 0$ an $M = M(\epsilon, t)$ such that

(9) $$| \rho_m(\pm t) - \alpha | < \epsilon \text{ for every } m \geqq M(\epsilon, t).$$

Furthermore, since $\rho_m \subset \| \phi_n \|$ for $m = 1, 2, \cdots$, there exist constants c_n^m $(m, n = 1, 2, \cdots)$ such that

$$\phi_n(x - c_n^m) \to \rho_m(x) \text{ as } n \to \infty \qquad (m = 1, 2, \cdots).$$

Hence if $t > 0$ is such that neither $x = t$ or $x = -t$ is contained in the at most enumerable set which consists of the points x at which at least one ρ_m is discontinuous, then, by the definition of the symbol $\to$, one can choose an $N = N(\epsilon, m, t)$ such that

$$| \phi_n(\pm t - c_n^m) - \rho_m(\pm t) | < \epsilon \text{ for every } n \geqq N(\epsilon, m, t);$$
$$(m = 1, 2, \cdots).$$

Consequently, from (9),

$$(10) \quad | \phi_n(\pm t - c_n^m) - \alpha | < 2\epsilon, \text{ if } m \geqq M(\epsilon, t), \ n \geqq N(\epsilon, m, t).$$

Since $\phi_n(x)$ is monotone and α is independent of x, it is clear that (10) remains valid if one replaces $\phi_n(\pm t - c_n^m)$ by $\phi_n(x - c_n^m)$, where x is any number between $-t$ and t. Hence

$$(11) \quad | \phi_n(x - c_n^m) - \alpha | < 2t^{-1}, \text{ if } | x | \leqq t, \ m \geqq M(t), \ n \geqq N(t),$$

where

$$M(t) = M(t^{-1}, t), \ N(t) = N(t^{-1}, M(t), t); \qquad (\epsilon = t^{-1}).$$

One can clearly assume that

$$N(t') < N(t''), \text{ if } t' < t''.$$

Since t is any positive number not belonging to an at most enumerable set, one can choose $t = t_1, t_2, \cdots$, where $t_k \to \infty$ as $k \to \infty$. For a fixed n which is not less $N(t_1)$, let the integer $k = k_n$ be defined by the condition

$$N(t_{k_n}) \leqq n < N(t_{k_{n+1}}).$$

Then

$$(12) \qquad k_n \to \infty \text{ and } t_{k_n} \to \infty \text{ as } n \to \infty.$$

Now, on placing

$$d_n = c_n^m, \text{ where } m = M(t_{k_n}),$$

it is clear from (11) that

$$| \phi_n(x - d_n) - \alpha | < 2t_{k_n}^{-1}, \text{ if } | x | \leqq t_{k_n}, \ n \geqq N(t_{k_n}).$$

Hence (12) implies that

$$\phi_n(x - d_n) \to \alpha, \qquad n \to \infty,$$

for every x. Thus $\alpha \subset \| \phi_n \|$. This completes the proof of Lemma 2.

(VI) If $\rho(x)$ is contained in $\| \phi_n \|$, then so are the constant functions $\alpha = \rho(-\infty)$ and $\alpha = \rho(+\infty)$.

This is clear from Lemma 2, since if $\rho_m(x) = \rho(x \pm m)$, then $\rho_m \subset \| \phi_n \|$, by (6).

Distribution functions. A monotone non-decreasing function $\phi(x)$ is said to be a distribution function if the set $]\phi[$, defined in (I), is the interval $0 < y < 1$, i. e., if $\phi(-\infty) = 0$ and $\phi(+\infty) = 1$. By the spectrum of a distribution function $\phi(x)$ is meant the set of those points $x = x_0$ for which $\phi(x_0 - \epsilon) \neq \phi(x_0 + \epsilon)$ whenever $\epsilon \neq 0$.

The metric defined under (II), when applied to the space of all distribution functions, defines a topology which is equivalent to the one defined by the symbol $\phi_n \rightarrow \phi$. In fact, since (3_2) and (3_3) are satisfied if ϕ_n and ϕ are distribution functions, (III) clearly implies

(VII) If $\phi_1, \phi_2, \cdots$ and ϕ are distribution functions, then $\phi_n \rightarrow \phi$ if and only if $| \phi_n ; \phi | \rightarrow 0$.

An obvious corollary of (VII) is the well-known fact that if ϕ_n and ϕ are distribution functions and ϕ is continuous for $-\infty < x < +\infty$, then ϕ_n cannot tend to ϕ unless the convergence is uniform for $-\infty < x < +\infty$. It is clear from (2) that (VII) implies

(VIII) If $\phi_1, \phi_2, \cdots$ are distribution functions, then there exists a distribution function ϕ satisfying $\phi_n \rightarrow \phi$ if and only if $\lim\limits_{\substack{n=\infty \\ m=\infty}} | \phi_n ; \phi_m | = 0$.

Throughout the paper, use will be made of the following notation:

(IX) If $\chi(x)$ is a monotone function for which

$$0 \leqq \chi(-\infty) \leqq \chi(+\infty) \leqq 1,$$

let there be defined for every positive number t a distribution function $[\chi]_t$ by placing

$$[\chi(x)]_t = 0, \text{ if } -\infty < x < -t;$$
$$[\chi(x)]_t = \chi(x), \text{ if } -t \leqq x \leqq t;$$
$$[\chi(x)]_t = 1, \text{ if } t < x < +\infty.$$

Thus $[\chi(x)]_t$ is a distribution function which is defined for every distribution function $\chi(x)$ and for certain functions $\chi(x)$ which are not distribution functions.

(X) If $\{\psi_m\}$ is a sequence of distribution functions and ρ a monotone non-decreasing function which need not be a distribution function, then $\psi_m \to \rho$ if and only if $|[\psi_m]_t; [\rho]_t| \to 0$ for every fixed $t > 0$.

This is clear from (VII) and (IX).

The set $\| \phi_n \|$ in case of distribution functions ϕ_n. If $\phi, \phi_2, \cdots$ are distribution functions, it is easily seen from (IV) that $\| \phi_n \|$ contains the constant functions $\alpha = 0$ and $\alpha = 1$. It is shown by the example

$$\phi_{2n}(x) = \tfrac{1}{2}(1 + \operatorname{sgn} x), \qquad \phi_{2n+1}(x) = \tfrac{1}{2}(1 + \tfrac{1}{\pi} \operatorname{arc\,tan} x)$$

that the set $\| \phi_n \|$ belonging to a sequence $\{\phi_n\}$ of distribution functions need not contain any function distinct from the constant functions $\alpha = 0$ and $\alpha = 1$. In particular, the set $\| \phi_n \|$ belonging to a sequence $\{\phi_n\}$ of distribution functions need not contain a distribution function.

THEOREM 1. *Every sequence $\{\phi_n\}$ of distribution functions contains a subsequence $\{\psi_m\}$ which has the following properties:*

(i) *There exists an at most enumerable set of mutually disjoint open subintervals $\alpha_k < y < \beta_k$ of $0 \leqq y \leqq 1$ such that a constant function is contained in $\| \psi_m \|$ if and only if its value y is not contained in any of these intervals $\alpha_k < y < \beta_k$.*

(ii) *There exists for each of these intervals $\alpha_k < y < \beta_k$ a monotone non-decreasing function $\rho_k(x)$ such that the set $]\rho_k[$, defined in* (I), *is the interval $\alpha_k < y < \beta_k$ and a non-constant function is contained in $\| \psi_m \|$ if and only if it is congruent with $\rho_k(x)$.*

The example mentioned before shows that Theorem 1 becomes false if one replaces a suitably chosen subsequence $\{\psi_m\}$ of $\{\phi_n\}$ by $\{\phi_n\}$ itself. The proof of Theorem 1 will be based on the following facts (XI), (XII), (XIII):

(XI) If $\{\phi_n\}$ is a sequence of distribution functions and s a given number such that $0 \leqq s \leqq 1$, then $\{\phi_n\}$ contains a subsequence $\{\psi_m\}$ such that for some element $\rho^s = \rho^s(x)$ of the function set $\| \psi_m \|$ the number s is contained in $]\rho^s[$.

Proof of (XI). The assumptions of (XI) clearly imply that

$$\phi_n(b_n - 0) \leqq s \leqq \phi_n(b_n + 0) \tag{13}$$

for every n and for some number $b = b_n$. Since $\{\phi_n(x + b_n)\}$ is a sequence of distribution functions, it contains a convergent subsequence. Let $\{\psi_m(x)\}$

and $\{a_m\}$ be the corresponding subsequences of $\{\phi_n(x)\}$ and $\{-b_n\}$ respectively and let $\sigma(x)$ denote the limit function of $\{\psi_m(x-a_m)\}$. Since $\sigma(x)$ is contained in $\| \psi_m \|$, so are, by (VI), the constant functions $\sigma(\pm \infty)$. Hence, on placing $\rho^s(x) = \sigma(x)$ or $\rho^s(x) = \sigma(\pm \infty)$ according as the number s is contained in $]\sigma[$ or is equal to $\sigma(\pm \infty)$, the statement of (XI) follows.

(XII) Every sequence $\{\phi_n\}$ of distribution functions contains a subsequence $\{\psi_m\}$ such that there exists for every rational number r, where $0 \leqq r \leqq 1$, an element ρ^r of $\| \psi_m \|$ for which r is contained in $]\rho^r[$. The

Proof of (XII) follows from (XI) by a straight-forward application of the diagonal principle.

(XIII) If a sequence $\{\psi_m\}$ of distribution functions is such that $\| \psi_m \|$ contains for every rational number r, where $0 \leqq r \leqq 1$, an element ρ^r for which $]\rho^r[$ contains r, then $\| \psi_m \|$ contains for every real number y, where $0 \leqq y \leqq 1$, an element ρ^y for which $]\rho^y[$ contains y.

Proof of (XIII). In the proof that ρ^y exists for a given y, it may be assumed without loss of generality that y is neither contained in a $]\rho^{r}[$ nor is y a boundary point of a $]\rho^r[$; cf. (VI). It is clear that, under these assumptions,

$$\rho^{r_n}(x) \to y, \qquad n \to \infty, \qquad (-\infty < x < +\infty)$$

whenever the sequence $\{r_n\}$ of rational numbers is such that $r_n \to y$. It follows, therefore, from Lemma 2 that the constant function $\alpha = y$ is contained in $\| \phi_n \|$. Thus the requirement of (XIII) is satisfied by the constant function $\rho^y(x) \equiv y$.

Proof of Theorem 1. Let $\{\psi_m\}$ be a subsequence of $\{\phi_n\}$ such that $\| \psi_m \|$ has the property stated under (XII). Then there exists, by (XIII), for every y, where $0 \leqq y \leqq 1$, a $\rho^y \subset \| \psi_m \|$ such that y is contained in $]\rho^y[$. If u and v are two distinct y-values and neither ρ^u nor ρ^v is a constant function, then Lemma 1 implies that the two open intervals $]\rho^u[$, $]\rho^v[$ are either disjoint or coincident, and that in the latter case ρ^u and ρ^v are congruent. Consequently, there exists in the interval $0 \leqq y \leqq 1$ an at most enumerable set of mutually disjoint open intervals $\alpha_k < y < \beta_k$ such that an open y-interval is an interval $\alpha_k < y < \beta_k$ if and only if it is an interval $]\rho^y[$ belonging to some non-constant function ρ^y, where $0 \leqq y \leqq 1$. Hence it is clear from (XIII) that if a number y, where $0 \leqq y \leqq 1$, is in none of the intervals $\alpha_k < y < \beta_k$, then the

constant function $\alpha = y$ is contained in $\| \psi_m \|$. On combining this with Lemma 2, Theorem 1 follows.

It is understood that the open set formed by the intervals $\alpha_k < y < \beta_k$ of Theorem 1 can be the empty set.

(XIV) If a subsequence $\{\psi_m\}$ of a sequence $\{\phi_n\}$ of distribution functions satisfies the requirements (i), (ii) of Theorem 1, then $\| \psi_m \|$ either contains a distribution function or it contains a constant function α, where $0 < \alpha < 1$.

This is clear from Theorem 1.

(XV) If a sequence $\{\phi_n\}$ of distribution functions is such that not every constant function α $(0 \leqq \alpha \leqq 1)$ is contained in $\| \phi_n \|$, then some subsequence $\{\psi_m\}$ of $\{\phi_n\}$ is such that $\| \psi_m \|$ contains a non-constant function.

Proof. Suppose that there exists a constant function α $(0 \leqq \alpha \leqq 1)$ which is not contained in $\| \phi_n \|$. Then $\alpha \neq 0$ and $\alpha \neq 1$. Hence there exists a sequence $\{c_n\}$ of numbers such that

$$\phi_n(c_n - 0) \leqq \alpha \leqq \phi_n(c_n + 0). \tag{14}$$

Since the sequence of the distribution functions $\phi_n(x + c_n)$ cannot tend to the constant function α, it contains a subsequence which tends to a limit function $\rho(x) \not\equiv \alpha$. Hence it is clear from (14) that $\rho(x)$ is not a constant function. Finally, if $\{\psi_m(x)\}$ is that subsequence of $\{\phi_n(x)\}$ which corresponds to the subsequence of $\{\phi_n(x + c_n)\}$ defining $\rho(x)$, then $\rho \subset \| \psi_m \|$. This completes the proof of (XV).

THEOREM 2. *If a subsequence $\{\psi_k\}$ of a sequence $\{\phi_n\}$ of distribution functions satisfies the requirements of Theorem 1, then one can choose for every $\epsilon > 0$ and for every $t > 0$ an $M = M(\epsilon, t) > 0$ such that*

(i) *there exists for every $m > M(\epsilon, t)$ and for every function $\rho \subset \| \psi_k \|$ a number $c = c_m$ for which*

$$|[\psi_m(x - c)]_t;\ [\rho(x)]_t| < \epsilon; \tag{15}$$

(ii) *there exists for every $m > M(\epsilon, t)$ and every number c a ρ such that* (15) *is satisfied for this $\rho = \rho_m{}^c$.* It is understood that the symbols $|\,;\,|$ and

$[\ ]_t$ are those defined under (II) and (IX).

Proof. For a fixed $\epsilon > 0$, choose $K = K_\epsilon$ values y_j such that

$$0 = y_1 < \cdots < y_j < \cdots < y_K = 1 \text{ and } y_j - y_{j-1} < \tfrac{1}{2}\epsilon. \tag{16}$$

By Theorem 1, there exists for every j an element $\chi = \chi_j$ of $\| \psi_k \|$ such that the y-set $]\chi_j[$ contains the point $y = y_j$. Choose a number a_j such that

$$\chi_j(a_j - 0) \leqq y_j \leqq \chi_j(a_j + 0)$$

and put

$$\tau_j(x) = \chi_j(x + a_j).$$

Then

$$\tau_j(-0) \leqq y_j \leqq \tau_j(+0), \tag{17}$$

and $\tau_j(x)$ is, by (6), contained in $\| \psi_k \|$. It follows, therefore, from (X) that one can choose for every j and for every $t > 0$ an $N = N_j(\epsilon, t)$ such that

$$|[\psi_m(x - c_j{}^m)]_{2t};\ [\tau_j(x)]_{2t}| < \tfrac{1}{2}\epsilon \text{ for every } m > N_j(\epsilon, t), \tag{18}$$

if the number $c_j{}^m$ is suitably chosen. Notice that $c_j{}^m$ can be chosen as independent of ϵ. On placing

$$M(\epsilon, t) = \operatorname{Max}\big(N_1(\epsilon, t), \cdots, N_j(\epsilon, t), \cdots, N_K(\epsilon, t)\big), \text{ where } K = K_\epsilon, \tag{19}$$

the statement (i) of Theorem 2 may be proved as follows:

Let ρ be a given element of $\| \psi_k \|$. If $\rho(x)$ is of the form $\tau_j(x - b)$, where $j = 1, \cdots, K_\epsilon$ and b is a number between $-t$ and t, then (i) is clear from (18) and (19). Suppose, therefore, that $\rho(x)$ is for no j of the form $\tau_j(x - b)$, where $| b | \leqq t$. Then, on the one hand, there exists by Lemma 1 a j such that

$$y_j \leqq \rho(x) \leqq y_{j+1} \text{ for } -t \leqq x \leqq t,$$

and, on the other hand, for this j,

$$y_j \leqq \tau_j(x + t) \leqq \rho(x) \text{ for } -t \leqq x \leqq t,$$

as seen from Lemma 1 and from (17). It follows, therefore, from (16) that, for this j,

$$|[\tau_j(x + t)]_t;\ [\rho(x)]_t| < \tfrac{1}{2}\epsilon;$$

cf. (II) and (IX). Since, by (18) and (19),

$$|[\psi_m(x + t - c_j{}^m)]_t;\ [\tau_j(x + t)]_t| < \tfrac{1}{2}\epsilon \text{ for } m > M(\epsilon, t),$$

it is seen from (1_3) that

$$|[\psi_m(x + t - c_j{}^m)]_t;\ [\rho(x)]_t| < \tfrac{1}{2}\epsilon + \tfrac{1}{2}\epsilon \text{ for } m > M(\epsilon, t).$$

Hence, (i) in Theorem 2 is satisfied by $c = c_j{}^m - t$.

The converse statement of Theorem 2, namely (ii), is similarly proved.

Convolutions. If ϕ_1, ϕ_2 are two distribution functions, there exists a unique distribution function $\phi_1 * \phi_2$ which is defined by

$$(20) \qquad \phi_1(x) * \phi_2(x) = \int_{-\infty}^{+\infty} \phi_1(x-u)\, d\phi_2(u)$$

and is called the convolution of $\phi_1(x)$ and $\phi_2(x)$. It is known that

$$\phi_1 * \phi_2 = \phi_2 * \phi_1 \text{ and } (\phi_1 * \phi_2) * \phi_3 = \phi_1 * (\phi_2 * \phi_3).$$

It is seen from (20) that, for arbitrary numbers c_1, c_2,

$$(21) \quad \phi_1(x-c_1) * \phi_2(x-c_2) = \phi(x-c_1-c_2), \text{ where } \phi(x) = \phi_1(x) * \phi_2(x).$$

Use will be made also of the following facts:[8]

(XVI) If $\phi_n^1(x), \phi_n^2(x), \phi^1(x), \phi^2(x)$ are distribution functions such that $\phi_n^1 \to \phi^1$ and $\phi_n^2 \to \phi^2$, then $\phi_n^1 * \phi_n^2 \to \phi^1 * \phi^2$.

(XVII) If $\phi_n^1(x)$, $\phi_n^2(x)$, $\phi(x)$ are distribution functions such that $\phi_n^1 \to \phi$ and $\phi_n^1 * \phi_n^2 \to \phi$, then $\phi_n^2 \to \omega$, where $\omega(x) = \frac{1}{2}(1 + \operatorname{sgn} x)$.

A fact similar to (XVII) is

(XVIII) If $\phi_n^1(x)$, $\phi_n^2(x)$ are distribution functions such that $\phi_n^1 * \phi_n^2 \to \omega$, where $\omega(x) = \frac{1}{2}(1 + \operatorname{sgn} x)$, then there exists a sequence $\{c_n\}$ of numbers such that $\phi_n^1(x + c_n) \to \omega(x)$ and $\phi_n^2(x - c_n) \to \omega(x)$.

Proof of (XVIII). On placing $\omega_n = \phi_n^1 * \phi_n^2$, the assumption of (XVIII) is that, for every $\epsilon > 0$ and for some $N = N(\epsilon)$,

$$1 - \epsilon < \omega_n(\epsilon) - \omega_n(-\epsilon), \text{ if } n \geqq N(\epsilon).$$

Since $\omega_n = \phi_n^1 * \phi_n^2$,

$$\omega_n(\epsilon) - \omega_n(-\epsilon) = \int_{-\infty}^{+\infty} [\phi_n^1(\epsilon - u) - \phi_n^1(-\epsilon - u)]\, d\phi_n^2(u),$$

and so one can assume that there exists a u for which

$$\omega_n(\epsilon) - \omega_n(-\epsilon) \leqq \phi_n^1(\epsilon - u) - \phi_n^1(-\epsilon - u).$$

Consequently, if c_ϵ^n denotes this u,

$$1 - \epsilon < \phi_n^1(\epsilon - c_\epsilon^n) - \phi_n^1(-\epsilon - c_\epsilon^n), \text{ if } n \geqq N(\epsilon).$$

[8] B. Jessen and A. Wintner, *loc. cit.* [4], Section 3.

One can assume that $N(1/k) < N(1/(k+1))$, where $k = 1, 2, \cdots$. For a given $n > N(1)$, define an integer $k = k_n$ by the requirement that

$$N(1/k_n) < n \leqq N(1/k_{n+1}).$$

Then $k_n \to \infty$ as $n \to \infty$ and, for every n,

$$1 - k_n^{-1} < \phi_n^1(c_n + k_n^{-1}) - \phi_n^1(c_n - k_n^{-1}),$$

where c_n denotes the negative value of c_ϵ^n for $\epsilon = k_n^{-1}$. Since $k_n^{-1} \to 0$ as $n \to \infty$, it follows that

$$\phi_n^1(x + c_n) \to \tfrac{1}{2}(1 + \operatorname{sgn} x) \text{ as } n \to \infty.$$

This, when combined with (21), clearly completes the proof of (XVIII).

A lemma on convolutions. The object of this section is the proof of a somewhat involved fact which isolates an essential part in the proof of Theorem 3 of the next section. The lemma in question (Lemma 3) is to the effect that if a convolution process $\phi_1 * \cdot$, when applied to a ϕ_2, flattens ϕ_2 strongly in the large, then also the local flattening of ϕ_2 must be quite strong. Certain weaker results to the same effect have been indicated by Lévy.[5]

LEMMA 3. *Let $\lambda(x)$, $\mu(x)$ be two distribution functions such that there exist four positive constants p, q, t, s which have the following properties:*

(i) $6q < p(1-p)$ *and* $p < 1$, *hence* $q < 1$;

(ii) $\lambda(x + 2t) - \lambda(x) \leqq p + q$ *for every* x;

(iii) $\lambda(s) - \lambda(-s) \geqq 1 - q$;

(iv) $\mu(t) - \mu(-t) \geqq p - 2q$;

(v) $\mu(t + 2s) - \mu(-t - 2s) \leqq p + q$.

Then there cannot exist a distribution function $\nu(x)$ such that $\lambda(x) * \nu(x) = \mu(x)$.

Proof. Suppose, if possible, that there exists a distribution function $\nu(x)$ such that

$$\mu(x) = \int_{-\infty}^{+\infty} \lambda(x - u)\, d\nu(u).$$

Then, from (iv),

$$p - 2q \leqq \left\{ \int_{-\infty}^{-s-t} + \int_{-s-t}^{s+t} + \int_{s+t}^{+\infty} \right\} [\lambda(t-u) - \lambda(-t-u)]\, d\nu(u),$$

where, according to (ii),

$$\int_{-s-t}^{s+t} [\ ]d\nu(u) \leqq (p+q)[\nu(s+t)-\nu(-s-t)],$$

while, according to (iii),

$$\left\{\int_{-\infty}^{-s-t} + \int_{s+t}^{+\infty}\right\}[\ ]d\nu(u) \leqq q\int_{-\infty}^{+\infty} d\nu(u) = q,$$

so that

$$p-2q \leqq q+(p+q)[\nu(s+t)-\nu(-s-t)].$$

On the other hand, from (v) and from the assumption $\mu = \lambda * \nu$,

$$p+q \geqq \mu(t+2s)-\mu(-t-2s)$$
$$= \int_{-\infty}^{+\infty} [\lambda(2s+t-u)-\lambda(-2s-t-u)]d\nu(u),$$

where $\int_{-\infty}^{+\infty} \geqq \int_{s+t}^{-s-t}$, and so, $\lambda(x)$ being a non-decreasing function which satisfies (iii),

$$p+q \geqq (1-q)[\nu(s+t)-\nu(-s-t)].$$

Consequently, by the inequality found before for $p-2q$,

$$p-2q \leqq q+(p+q)[(p+q)/(1-q)],$$

and so

$$p-p^2 \leqq 3q(1+p).$$

Since this contradicts (i), the proof of Lemma 3 is complete.

Convolution sequences. A sequence $\{\phi_n\}$ of distribution functions will be called a convolution sequence if there exists a sequence $\{\sigma_n\}$ of distribution functions such that

$$(22) \qquad \phi_n = \sigma_1 * \cdots * \sigma_n, \text{ i. e., } \phi_n = \phi_{n-1} * \sigma_n \qquad (\phi_1 = \sigma_1,\ n = 2, 3, \cdots).$$

(XIX) Every subsequence of a convolution sequence is a convolution sequence.

This is clear from $\chi_1 * (\chi_2 * \chi_3) = (\chi_1 * \chi_2) * \chi_3$.

(XX) If $\{\phi_n(x)\}$ is a convolution sequence, then so is $\{\phi_n(x - c_n)\}$ for every sequence of numbers c_n.

This is clear from (21).

THEOREM 3. *If, for a given convolution sequence* $\{\phi_n\}$, *one denotes by* $\| \phi_n \|_0$ *the function set obtained from* $\| \phi_n \|$ *by omitting the two constant functions* $\alpha = 0$, $\alpha = 1$, *then either every element of* $\| \phi_n \|$ *is a constant function or every element of* $\| \phi_n \|_0$ *is a distribution function. In the first case every constant function* α, *where* $0 \leqq \alpha \leqq 1$, *is an element of* $\| \phi_n \|$. *In the second case there exists a distribution function* ρ *such that a function is an element of* $\| \phi_n \|_0$ *if and only if it is congruent with* ρ. In the first case the convolution sequence $\{\phi_n\}$ will be said to be *flat*, in the second case *non-flat*.

The proof is somewhat lengthy and will be decomposed into the following steps (XXI_1), (XXI_2):

(XXI_1) If $\{\phi_n\}$ is a convolution sequence, then every non-constant function contained in the function set $\| \phi_n \|$ is a distribution function.

Proof of (XXI_1). Let $\{\psi_m\}$ be a subsequence of $\{\phi_n\}$ such that $\{\psi_m\}$ has the properties described in Theorem 1. Thus there exists a number $p \geqq 0$ such that every $\rho \subset \| \psi_m \|$ has, for $-\infty < x < +\infty$, a total variation not greater than p, while the total variation of some $\rho_0 \subset \| \psi_m \|$ is equal to p. It is seen from (5) that (XXI_1) will be proved if one shows that every non-constant function contained in $\| \psi_m \|$ is a distribution function. Suppose, if possible, that there exists in $\| \psi_m \|$ a non-constant function which is not a distribution function. Then, by the definition of ρ_0,

$$0 < p < 1, \text{ where } p = \rho_0(+\infty) - \rho_0(-\infty). \tag{23}$$

Hence one can choose a $q > 0$ for which condition (i) of Lemma 3 is satisfied. It is also seen from (23) that, for some $t > 0$,

$$\rho_0(t) - \rho_0(-t) > p - q. \tag{24}$$

On the other hand, by the definition of p,

$$p \geqq \rho(+\infty) - \rho(-\infty) \geqq \rho(x + 2t) - \rho(x)$$

for every $\rho \subset \| \psi_m \|$ and for every x. Hence (i) of Theorem 2 assures the existence of a $\psi_k \subset \{\psi_m\}$ such that

$$\psi_k(x + 2t) - \psi_k(x) \leqq p + q \text{ for every } x.$$

On denoting this ψ_k by λ, condition (ii) of Lemma 3 is satisfied. Since λ is a distribution function and $q > 0$, there clearly exists a number $s > 0$ which satisfies condition (iii) of Lemma 3. Since $\rho_0 \subset \| \psi_m \|$, there exists a sequence $\{c_m\}$ of numbers such that, at every continuity point x of ρ_0,

$$\psi_m(x - c_m) \to \rho_0(x), \qquad m \to \infty.$$

Hence it is seen from (24) and (23) that there exists[9] a $\psi_l \subset \{\psi_m\}$ such that $l > k$ and

$$\psi_l(t - c_l) - \psi_l(-t - c_l) > p - q - q,$$
$$\psi_l(t + 2s - c_l) - \psi_l(-t - 2s - c_l) < p + q.$$

On denoting the distribution function $\psi_l(x - c_l)$ by $\mu(x)$, conditions (iv) and (v) of Lemma 3 are satisfied. Since $\{\psi_m\}$ is, by (XIX), a convolution sequence, and since

$$\lambda(x) = \psi_k(x), \ \mu(x) = \psi_l(x - c_l) \ \text{ and } \ l > k,$$

there exists, by (20) and (21), a distribution function $\nu(x)$ such that $\lambda * \nu = \mu$. Since this contradicts Lemma 3, the proof of (XXI_1) is complete.

(XXI_2) If a distribution function ρ is contained in $\| \psi_m \|$, where $\{\psi_m\}$ is a subsequence of a convolution sequence $\{\phi_n\}$, then ρ is contained in $\| \phi_n \|$.

Proof of (XXI_2). Suppose, if possible, that (XXI_2) is false. Then (XXI_1) implies that the set $\| \chi_k \|_0$ belonging to a suitably chosen subsequence $\{\chi_k\}$ of $\{\phi_n\}$ contains a function π which is not congruent with ρ. In particular,

$$(25) \qquad \tau_k \to \pi, \qquad k \to \infty,$$

holds for a sequence of distribution functions $\tau_k(x)$ of the form $\chi_k(x - c_k)$. Since $\psi_m(x)$ and $\tau_k(x)$ are of the form $\phi_n(x - a_n)$, and since $\{\phi_n\}$ is a convolution sequence, there exist, by (21), for every m two distribution functions λ_m, μ_m and two positive integers k, j such that

$$(26) \qquad \psi_m * \lambda_m = \tau_k, \text{ where } k \to \infty \text{ as } m \to \infty,$$

and

$$(27) \qquad \tau_k * \mu_m = \psi_j, \text{ where } j > m.$$

Hence

$$(28) \qquad \psi_m * \lambda_m * \mu_m = \psi_j \text{ for some } j > m.$$

[9] The discontinuity points of ρ_0, if any, do not interfere with the possibility of this conclusion, since they form an at most enumerable set.

Since $\rho \subset \| \psi_m \|$, one can assume without loss of generality [cf. (IV) and (21)] that $\psi_m \to \rho$ as $m \to \infty$. Then (28) implies, in view of (XVII), that $\lambda_m * \mu_m \to \omega$. It follows, therefore, from (XVIII) that, for a suitably chosen sequence of numbers b_m,

$$(29) \qquad \lambda_m(x - b_m) \to \omega(x) \text{ as } m \to \infty.$$

Since, from (26) and (25),

$$(30) \qquad \psi_m(x) * \lambda_m(x) \to \pi(x),$$

and since $\pi(x)$ is not one of the constant functions equal to 0 or 1, it is easily inferred from (XVI) and (31) that the numbers b_m tend to a limit b as $m \to \infty$. Hence (29) can be replaced by

$$(31) \qquad \lambda_m(x) \to \omega(x + b).$$

It follows, therefore, from (30) (XVI) and from the definition $\psi_m(x) \to \rho(x)$ of ρ that

$$(32) \qquad \rho(x) * \omega(x + b) = \pi(x).$$

This means in view of the definition of $\omega(x)$ that $\rho(x + b) = \pi(x)$, i. e., that π is a distribution function and that ρ and π are congruent. Since this contradicts the assumption, the proof of (XXI_2) is complete.

Proof of Theorem 3. Suppose that $\| \phi_n \|$ does not contain all constant functions α, where $0 \leqq \alpha \leqq 1$. Then $\{\phi_n\}$ contains, by (XV), a subsequence $\{\psi_m\}$ such that $\| \psi_m \|$ contains a non-constant function, say ρ. This ρ is, by (XXI_1), a distribution function. Furthermore, $\rho \subset \| \phi_n \|$, by ($XXI_2$). Finally, it is seen from (6) and Lemma 1 that a function is contained in $\| \phi_n \|_0$ if and only if it is congruent with ρ. This completes the proof of Theorem 3.

There arises the question how to decide whether or not a given convolution sequence $\{\phi_n\}$ is flat in the sense of Theorem 3. In order to obtain criteria to this effect, put

$$(33_1) \qquad E(\chi) = \int_{-\infty}^{+\infty} x d\chi(x),$$

if the distribution function χ has a first moment, which is certainly the case if χ has a finite second moment

$$(33_2) \qquad F(\chi) = \int_{-\infty}^{+\infty} x^2 d\chi(x).$$

In the latter case, let

$$(33) \qquad D(\chi) = F(\chi) - [E(\chi)]^2.$$

Thus $D(\chi) = \int_{-\infty}^{+\infty} [x - E(\chi)]^2 d\chi(x)$; hence

$$(34) \qquad D(\chi) \geqq 0,$$

where $D(\chi) = 0$ if and only if χ is congruent with $\omega = \frac{1}{2}(1 + \operatorname{sgn} x)$. If $F(\chi) = +\infty$, put $D(\chi) = +\infty$.

It is clear from (33_1) that

$$(35_1) \qquad E(\chi_1) = E(\chi_2) + c, \text{ if } \chi_1(x) = \chi_2(x - c).$$

Similarly, from (33_2),

$$(35_2) \qquad F(\chi_1) = F(\chi_2) + 2cE(\chi_2) + c^2, \text{ if } \chi_1(x) = \chi_2(x - c).$$

Consequently, from (33),

$$(35) \qquad D(\chi_1) = D(\chi_2), \text{ if } \chi_1 \text{ and } \chi_2 \text{ are congruent.}$$

Hence, on combining (IV), (21) and Theorem 3 with known[10] criteria for the convergence of infinite convolutions, it is seen that Theorem 3 can be completed by

THEOREM 4. *In order that a convolution sequence $\{\phi_n\}$ be non-flat, it is sufficient that, on using the notations* (33) *and* (22),

$$(36) \qquad \sum_{n=1}^{\infty} D(\sigma_m) < +\infty \qquad [\text{cf. } (34)].$$

This sufficient condition is necessary as well in case there exists a sufficiently large $L > 0$ such that the spectrum of every σ_n is a subset of an x-interval of suitably chosen position and of length L, where L is independent of n.

Convergent infinite convolutions. Let $\{\phi_n\}$ be a convolution sequence, i. e., a sequence of distribution functions which can be represented in the form (22). It is clear that $\{\phi_n\}$ can converge, in the sense of (3_1), to a ϕ which is not a distribution function. The infinite convolution $\phi = \sigma_1 * \sigma_2 * \cdots$ is said to be a convergent infinite convolution only if (22) satisfies (3_1) with a function ϕ which is a distribution function. On using this terminology, Theorem 3 clearly implies

---

[10] B. Jessen and A. Wintner, *loc. cit.* [4], Theorem 4 and Theorem 5.

THEOREM 5. *If a convolution sequence*

$$\{\phi_n\} = \{\sigma_1 * \cdots * \sigma_n\} \tag{37}$$

tends, in the sense of (3_1), *to a limit function* ϕ, *and if the infinite convolution* $\phi = \sigma_1 * \sigma_2 * \cdots$ *is not convergent, then* ϕ *is a constant function.*

Theorem 3 also implies [11]

THEOREM 6. *There exists for every non-flat convolution sequence* (37) *a sequence* $\{c_n\}$ *of numbers for which the infinite convolution*

$$\sigma_1(x - c_1) * \sigma_2(x - c_2) * \cdots \tag{38}$$

is convergent. If $\{c_n{}^1\}$ *and* $\{c_n{}^2\}$ *are two such sequences* $\{c_n\}$, *then the two corresponding infinite convolutions* (38) *represent congruent distribution functions.*

If, on the other hand, a convolution sequence (37) *is flat, then the infinite convolution* (38) *is divergent for every* $\{c_n\}$.

In what follows, use will be made of the following fact which is merely a restatement [12] of a part of the fundamental result of Khintchine and Kolmogoroff [2] concerning "equivalent" series of independent random variables:

(XXII) The infinite convolution $\sigma_1 * \sigma_2 * \cdots$ is convergent if and only if so is the infinite convolution $[\sigma_1]_t * [\sigma_2]_t * \cdots$ for a $t > 0$ (in which case the same holds for every $t > 0$). It is understood that $[\sigma]_t$ denotes the distribution function defined in (IX).

On combining (XXII) with the known convergence criterion [10] for a convolution of distribution functions with uniformly bounded spectra, one obtains

(XXIII) The infinite convolution $\sigma_1 * \sigma_2 * \cdots$ is convergent if and only if so are both series

$$\sum_{n=1}^{\infty} E([\sigma_n]_t), \qquad \sum_{n=1}^{\infty} D([\sigma_n]_t) \qquad \text{[cf. } (33_1), (33)]$$

for a $t > 0$ (in which case the same holds for every $t > 0$).

Absolutely convergent infinite convolutions. Two infinite convolutions, $\sigma_1' * \sigma_2' * \cdots$ and $\sigma_1'' * \sigma_2'' * \cdots$, will be said to be rearrangements of each

[11] This theorem has been stated without a detailed proof by P. Lévy, *loc. cit.* [5], p. 340; cf. also p. 337.

[12] Cf. B. Jessen and A. Wintner, *loc. cit.* [4], Theorems 32 and 34.

other if the sequences $\{\sigma_n'\}$ and $\{\sigma_n''\}$ are permutations of each other. An infinite convolution $\sigma_1 * \sigma_2 * \cdots$ is said to be absolutely convergent if every rearrangement of it is a convergent infinite convolution. It is known [8] that in this case all rearrangements represent the same distribution function. The definition of an absolutely convergent infinite convolution clearly implies

(XXIV) Both (XXII) and (XXIII) remain valid if one reads "absolutely convergent" instead of "convergent."

THEOREM 7. *There exists for every convergent infinite convolution* $\sigma_1(x) * \sigma_2(x) * \cdots$ *a sequence* $\{c_n\}$ *of numbers such that the infinite convolution* $\sigma_1(x-c_1) * \sigma_2(x-c_2) * \cdots$ *is absolutely convergent.* Needless to say, another sequence, $\{\bar{c}_n\}$, of numbers has this property if and only if $|c_1-\bar{c}_1| + |c_2-\bar{c}_2| + \cdots$ is a convergent series.

Proof. Choose a fixed $t>0$, put

$$\rho_n(x) = [\sigma_n(x)]_{2t}, \tag{39}$$

where $[\ ]_{2t}$ is defined by (IX), and let

$$(40_1)\quad \tau_n(x) = \rho_n(x-c_n); \qquad (40_2)\quad \chi_n(x) = \sigma_n(x-c_n),$$

where, on using the notation (33_1), the number c_n is chosen as follows:

$$(41)\quad c_n = 0 \text{ or } c_n = E(\rho_n) \text{ according as } |E(\rho_n)| > t \text{ or } |E(\rho_n)| \leqq t.$$

Thus $|c_n| \leqq t$, and so (39), (40_1), (40_2) and (IX) imply that

$$[\chi_n]_t = [\tau_n]_t.$$

Hence (XXIV) and (40_2) show that the infinite convolution (38) is absolutely convergent if and only if so is the infinite convolution $\tau_1(x) * \tau_2(x) * \cdots$, i. e., if and only if so are both series

$$(42_1)\quad \sum_{n=1}^{\infty} E(\tau_n); \qquad (42_2)\quad \sum_{n=1}^{\infty} D(\tau_n).$$

Now (42_1) is absolutely convergent, since, on the one hand,

$$E(\tau_n) = E(\rho_n) - c_n, \text{ by } (35_1) \text{ and } (40_1),$$

and, on the other hand,

$$E(\rho_n) = c_n \text{ for every sufficiently large } n,$$

as seen from (41) and from the fact that $\sum_{n=1}^{\infty} E(\rho_n)$ is, by (XXIII), (39) and the assumption of Theorem 7, a convergent series. On using (35) instead of (35_1) and (34) instead of (41), it is similarly shown that (42_2) is absolutely convergent. Consequently, the infinite convolution (38) is absolutely convergent. This completes the proof of Theorem 7.

Theorem 7 implies, in view of Theorem 6,

THEOREM 8. *A convolution sequence* (37) *is non-flat if and only if there exists a sequence* $\{c_n\}$ *of numbers such that the infinite convolution* (38) *is absolutely convergent.*[11]

Theorem 8 and Theorem 3 imply

THEOREM 9. *If two sequences* $\{\sigma_n'\}$, $\{\sigma_n''\}$ *of distribution functions are permutations of each other, then the two function sets* $\| \phi_n' \|_0$, $\| \phi_n'' \|_0$, *where* $\phi_n' = \sigma_1' * \cdots * \sigma_n'$, $\phi_n'' = \sigma_1'' * \cdots * \sigma_n''$, *are identical function sets. In other words, either both convolution sequences* $\{\phi_n'\}$, $\{\phi_n''\}$ *are flat or both are non-flat, and in the latter case the two distribution functions* ρ', ρ'', *which* $\| \phi_n' \|_0$, $\| \phi_n'' \|_0$ *determine up to congruences, are congruent.*

Theorems 3 and 8 can be interpreted as describing the possible behavior of any divergent infinite convolution. Correspondingly, Theorem 9 describes what can happen to a non-absolutely convergent infinite convolution upon an arbitrary rearrangement of its "factors." In the non-flat case, Theorems 8 and 9 imply the more precise fact that *the infinite convolution behaves upon a rearrangement exactly the same way as a certain numerical series,* which is determined by the sequence of the "factors" up to an additive absolutely convergent numerical series.

THE JOHNS HOPKINS UNIVERSITY.

AN ANALOGUE OF JACOBI'S CONDITION FOR THE PROBLEM OF MAYER WITH VARIABLE END POINTS.*

By Thomas Freeman Cope.

Introduction. The problem of Mayer with variable end points, as stated by Bliss,[1] is the determination of the properties of an arc

$$\text{(E)} \qquad y_i = y_i(x), \qquad x_1 \leqq x \leqq x_2, \qquad (i = 1, \cdots, n),$$

which minimizes the first of a set of functions

$$f_\rho[x_1, x_2, y(x_1), y(x_2)], \qquad (\rho = 1, \cdots, r \leqq 2n + 2),$$

in the class of similar arcs which make $f_2, \cdots, f_r$ vanish and besides satisfy the differential equations

$$(1) \qquad \phi_\alpha(x, y, y') = 0, \qquad (\alpha = 1, \cdots, m < n),$$

where $y(x)$ is a symbol for $y_1 \cdots y_n$. Bliss has shown [2] that if Ω denotes the function

$$\Omega = \lambda_1\phi_1 + \cdots + \lambda_m\phi_m$$

then a necessary condition for a minimum is that there shall exist m functions $\lambda_1(x), \cdots, \lambda_m(x)$, not all identically zero on x_1x_2, satisfying the differential equations

$$(2) \qquad \Omega_{y_i} - \frac{d}{dx}\Omega_{y'_i} = 0, \qquad (i = 1, \cdots, n),$$

and making all determinants of order $r + 1$ of the matrix

$$(3) \qquad \begin{Vmatrix} f_{\rho x_1} & f_{\rho y_{i1}} & f_{\rho x_2} & f_{\rho y_{i2}} \\ \Omega(x_1) - \Omega_{y'_i}(x_1)y'_{i1}, & \Omega_{y'_i}(x_1), & -\Omega(x_2) + \Omega_{y'_i}(x_2)y'_{i2}, & -\Omega_{y'_i}(x_2) \end{Vmatrix}$$

* Presented to the American Mathematical Society (Chicago), April, 1928. Received by the Editors March 1, 1937. This paper is a somewhat revised form of the author's dissertation, see 4. The great interest shown in the problem and methods of this paper, as attested by the bibliography at the end, seemed to the author to justify its publication at this time.

The numbers in the footnotes refer to the bibliography at the end where further references will be found.

[1] 1. [2] 1.

vanish. According to the usual convention of tensor analysis, it is understood here and elsewhere in this paper that a subscript i, j, α, β, etc., repeated in the same term indicates a sum. The subscripts x, y, y' denote partial derivatives. It is also understood that the arguments in f_ρ, Ω, and their derivatives are those belonging to E.

The preceding statement of the necessary condition is equivalent to the statement[3] that there must exist m functions $\lambda_\alpha(x)$, not all identically zero on x_1x_2, satisfying the equations (2), and r constants $l_1, \cdots, l_r$, not all zero, satisfying with them the $2n+2$ equations

$$(4) \quad f_{x_1} + f_{y_{i1}} y'_{i1} = f_{y_{i1}} - \Omega_{y'_i}(x_1) = -f_{x_2} - f_{y_{i2}} y'_{i2} = -f_{y_{i2}} - \Omega_{y'_i}(x_2) = 0,$$

where

$$f = l_1 f_1 + \cdots + l_r f_r.$$

In the first section of this paper, the hypotheses on which the analysis is based, are stated and preliminary notions and theorems considered. The first and second variations are computed in section 2. It is shown in section 3 that a minimum problem for the second variation may be formulated and moreover that it can be transformed into a problem of the same type as the original one. The differential equations and boundary conditions corresponding to (2) and (4) above for this auxiliary minimum problem are then given. In section 4 a boundary value problem associated with the second variation is stated and discussed, and by means of it a necessary condition for the original minimum problem is proved. This condition is essentially that for a minimizing arc for the original problem the boundary value problem of the second variation can have no solution for negative values of its parameter. It is then shown in section 5 that the boundary value problem of the second variation can be transformed into one that has been treated by Bliss. Section 6 is devoted to the task of proving that the transformed boundary value problem is "definitely self-adjoint," according to Bliss's definition. From this fact much information is automatically obtained about the characteristic constants and solutions of the original boundary value problem.

The form of the second variation appearing in section 2, the analogue of the Jacobi necessary condition of section 4, and the discussion of the boundary value problem of the second variation of sections 5 and 6 were first given, for the general problem of Mayer with variable end points, in my dissertation (see 4). Proofs of similar analogues have since been published by Morse (see 7, p. 524) and Reid (see 14, p. 840). The same authors have also

[3] 1, p. 311.

treated the boundary value problem of the second variation by methods quite different from mine and from one another (see 7, pp. 542-546, and 10). Analogues of the Jacobi necessary condition different in form from those mentioned above have been given by Bliss (see 8, p. 266) and Hestenes (see 12, p. 483).

1. Preliminary notions and theorems. The arc E is supposed to have the following properties: [4]

1. It is of class C''' and such that the functions ϕ_α, f_ρ are of class C^{IV} in a neighborhood R of the values (x, y, y') on E.

2. It satisfies the equations (1) and

$$f_\sigma = 0, \qquad (\sigma = 2, \cdots, r). \tag{5}$$

3. The matrix $\| \phi_{\alpha y'_i} \|$ has rank m at every point of E.

4. The matrix

$$\| f_{\rho x_1} f_{\rho y_{i1}} f_{\rho x_2} f_{\rho y_{i2}} \|, \qquad (\rho = 1, \cdots, r;\ i = 1, \cdots, n),$$

with $2n + 2$ columns and r rows is of rank r at the values of the arguments of the functions f_ρ on E.

Consider a one-parameter family of arcs [5]

$$\begin{aligned} y_i &= Y_i(x, \epsilon), && (i = 1, \cdots, n), \\ X_1(\epsilon) \leqq x &\leqq X_2(\epsilon), \quad X_\iota(0) = x_\iota, && (\iota = 1, 2), \end{aligned} \tag{6}$$

which contains E for $\epsilon = 0$ and satisfies the equations (1) and (5) for every ϵ in a neighborhood of $\epsilon = 0$. Its variations are by definition the expressions

$$\xi_\iota = X_{\iota\epsilon}(0), \qquad \eta_i(x) = Y_{i\epsilon}(x, 0).$$

The variations then satisfy the equations of variation

$$\Phi_\alpha(x, \eta, \eta') \equiv \phi_{\alpha y_i}\eta_i + \phi_{\alpha y'_i}\eta'_i = 0, \qquad (\alpha = 1, \cdots, m), \tag{7}$$

$$F_\sigma(\xi, \eta) = 0, \qquad (\sigma = 2, \cdots, r), \tag{8}$$

where

$$\begin{aligned} F_\rho(\xi, \eta) \equiv{} & (f_{\rho x_1} + f_{\rho y_{i1}} y'_{i1})\xi_1 + f_{\rho y_{i1}}\eta_{i1} \\ & + (f_{\rho x_2} + f_{\rho y_{i2}} y'_{i2})\xi_2 + f_{\rho y_{i2}}\eta_{i2}, \qquad (\rho = 1, \cdots, r). \end{aligned} \tag{9}$$

The functions y, y' occurring explicitly and in the derivatives are those defining E.

Consider now a system H of r sets of variations $\xi_\iota{}^\rho$, $\eta_i{}^\rho$ $(\rho = 1, \cdots, r;$

[4] 1, p. 307.

[5] 1, p. 307.

$\iota = 1, 2$), with η's of class C''' and satisfying the equations (7). Variations of this sort with ξ_1 and ξ_2 arbitrary constants are called *admissible* variations. A minimizing arc E is said to be normal for the problem under consideration when a system H of variations can be so selected that the matrix

$$\| F_\sigma(\xi^\rho, \eta^\rho) \|, \qquad (\sigma = 2, \cdots, r;\ \rho = 1, \cdots, r)$$

has rank $r - 1$. Let an admissible arc be defined as an arc of class C''' on x_1x_2, whose elements (x, y, y') all lie in R, and which satisfies the equations (1). The following theorem and its corollary with their proofs, which are omitted here, are similar to those given by Bliss in his lectures at Chicago in the summer of 1925.[6]

THEOREM. *For every normal minimizing arc E of class C''' on x_1x_2 for the Mayer problem with variable end points, there exists a one-parameter family of admissible arcs* (6) *containing E for $\epsilon = 0$ and satisfying the equations* (5). *The functions Y_i are of class C''' in x, and Y_i, Y'_i, X_ι $(i = 1, \cdots, n;\ \iota = 1, 2)$ of class C'' in ϵ, near $x_1 \leqq x \leqq x_2$, $\epsilon = 0$.*

COROLLARY. *If a set of admissible variations ξ, $\eta(x)$ for a normal minimizing arc E of class C''' on x_1x_2 for the Mayer problem with variable end points satisfies the equations* (8), *there exists a one-parameter family of admissible arcs* (6) *satisfying the end conditions* (5), *containing the arc E for $\epsilon = 0$, and having the set ξ, η as its variations along E. The functions Y_i are of class C''' in x and Y_i, Y'_i, X_ι of class C'' in ϵ near $x_1 \leqq x \leqq x_2$, $\epsilon = 0$.*

2. The first and second variations. Consider the minimizing arc E of the corollary. There must exist m functions $\lambda_\alpha(x)$ of class C', not all identically zero on x_1x_2, satisfying the equations (2), and r constants l_ρ, not all zero, satisfying the equations (4). Moreover, since E is normal, l_1 must be different from zero.[7] l_1 may then be chosen equal to unity, and in the following pages it will be assumed that such a choice of l_1 has been made.

Substitute the one-parameter family of the corollary in the functions f_ρ, ϕ_α and differentiate with respect to ϵ. The result is

$$\begin{aligned} \frac{df_1}{d\epsilon} &= F_1(X_\epsilon, Y_\epsilon), & \\ 0 &= F_\sigma(X_\epsilon, Y_\epsilon), & (\sigma = 2, \cdots, r), \\ 0 &= \Phi_\alpha(x, Y_\epsilon, Y'_\epsilon), & (\alpha = 1, \cdots, m), \end{aligned}$$

[6] 5, pp. 694-695, and 4, pp. 6-9.

[7] 1, p. 311.

where F_ρ and Φ_a with the arguments indicated are defined by equations (7) and (9), but with coefficients taken for $\epsilon = \epsilon$. Differentiating again with respect to ϵ and putting $\epsilon = 0$, we find

$$(10)\quad \begin{aligned} \left(\frac{d^2 f_1}{d\epsilon^2}\right)\Bigg|^{\epsilon=0} &= F_1(X_{\epsilon\epsilon}, Y_{\epsilon\epsilon}) + 2f_{1y_j}Y'_{j\epsilon}X_\epsilon\,|^1 + 2f_{1y_j}Y'_{j\epsilon}X_\epsilon\,|^2 + Q_1(X_\epsilon, Y_\epsilon),\\ 0 &= F_\sigma(X_{\epsilon\epsilon}, Y_{\epsilon\epsilon}) + 2f_{\sigma y_j}Y'_{j\epsilon}X_\epsilon\,|^1 + 2f_{\sigma y_j}Y'_{j\epsilon}X_\epsilon\,|^2 + Q_\sigma(X_\epsilon, Y_\epsilon),\\ 0 &= \Phi_a(x, Y_{\epsilon\epsilon}, Y'_{\epsilon\epsilon}) + 2\omega_a(x, Y_\epsilon, Y'_\epsilon),\\ &\qquad (j = 1, \cdots, n;\ \sigma = 2, \cdots, r;\ \alpha = 1, \cdots, m), \end{aligned}$$

where Q_1, Q_σ, $2\omega_a$ are quadratic forms in the arguments indicated, and where all the arguments are taken for $\epsilon = 0$. Multiply the first r equations of (10) by l_1, l_σ, respectively, and add. Then because of the equations (4) of section 1, the sum can be written

$$(11)\quad \left(\frac{d^2 f_1}{d\epsilon^2}\right)\Bigg|^{\epsilon=0} = \Omega_{y'_j}Y_{j\epsilon\epsilon}\Bigg|_{X_2(\epsilon)}^{X_1(\epsilon)} + 2\Omega_{y'_j}Y'_{j\epsilon}X_\epsilon\Bigg|_{X_2(\epsilon)}^{X_1(\epsilon)} + l_\rho Q_\rho(X_\epsilon, Y_\epsilon), \qquad (\rho = 1, \cdots, r),$$

where all the elements are taken for $\epsilon = 0$. Multiply now the last m equations of (10) by λ_a and add. The result is, if we put $\omega = \lambda_a\omega_a$,

$$\begin{aligned} 0 &= \Omega_{y_j}Y_{j\epsilon\epsilon} + \Omega_{y'_j}Y'_{j\epsilon\epsilon} + 2\omega(x, Y_\epsilon, Y'_\epsilon)\\ &= Y_{j\epsilon\epsilon}\left(\Omega_{y_j} - \frac{d}{dx}\Omega_{y'_j}\right) + \frac{d}{dx}\left(\Omega_{y'_j}Y_{j\epsilon\epsilon}\right) + 2\omega\\ &= \frac{d}{dx}\left(\Omega_{y'_j}Y_{j\epsilon\epsilon}\right) + 2\omega, \end{aligned}$$

since for $\epsilon = 0$, the Euler-Lagrange equations (2) are true. Integrating between $X_1(\epsilon)$ and $X_2(\epsilon)$ for $\epsilon = 0$, we obtain

$$(12)\qquad 0 = \Omega_{y'_j}Y_{j\epsilon\epsilon}\Bigg|_{X_1(\epsilon)}^{X_2(\epsilon)} + \int_{X_1(\epsilon)}^{X_2(\epsilon)} 2\omega(x, Y_\epsilon, Y'_\epsilon)\,dx.$$

Finally by multiplying the equations

$$0 = \Phi_a(x, Y_\epsilon, Y'_\epsilon)$$

by λ_a and adding, we find

$$0 = \Omega_{y_j}Y_{j\epsilon} + \Omega_{y'_j}Y'_{j\epsilon},$$

whence

$$(13)\qquad 0 = 2\Omega_{y_j}Y_{j\epsilon}X_\epsilon\Bigg|_{X_1(\epsilon)}^{X_2(\epsilon)} + 2\Omega_{y'_j}Y'_{j\epsilon}X_\epsilon\Bigg|_{X_1(\epsilon)}^{X_2(\epsilon)},$$

where as before we take $\epsilon = 0$. Now add (11), (12), and (13). The result will be the desired form of the second variation, after putting

$$X_{1\epsilon}(0) = \xi_1, \qquad X_{2\epsilon}(0) = \xi_2, \qquad Y_j(x, \epsilon)\,|^{\epsilon=0} = \eta_j(x),$$

namely,

$$(14) \qquad I_2 \equiv \left(\frac{d^2 f_1}{d\epsilon^2}\right)\Bigg|^{\epsilon=0} = \int_{x_1}^{x_2} 2\omega(x, \eta, \eta')\,dx + Q(\xi, \eta),$$

where the quadratic forms 2ω and Q are explicitly

$$(15) \qquad \begin{aligned}
2\omega &= P_{ij}\eta_i\eta_j + 2Q_{ij}\eta_i\eta'_j + R_{ij}\eta'_i\eta'_j;\\
Q &= A\xi_1^2 + 2B\xi_1\xi_2 + C\xi_2^2 + 2D_j\eta_j(x_1)\xi_1 + 2E_j\eta_j(x_2)\xi_1\\
&\quad + 2F_j\eta_j(x_1)\xi_2 + 2G_j\eta_j(x_2)\xi_2 + H_{ij}\eta_i(x_1)\eta_j(x_1)\\
&\quad + 2I_{ij}\eta_i(x_1)\eta_j(x_2) + J_{ij}\eta_i(x_2)\eta_j(x_2);\\
P_{ij} &= \Omega_{y_iy_j}, \qquad Q_{ij} = \Omega_{y_iy'_j}, \qquad R_{ij} = \Omega_{y'_iy'_j};\\
A &= \frac{d}{dx_1}\left(f_{x_1} + f_{y_{j1}}y'_j(x_1)\right), \qquad B = \frac{d}{dx_1}\left(f_{x_2} + f_{y_{j2}}y'_j(x_2)\right),\\
C &= \frac{d}{dx_2}\left(f_{x_2} + f_{y_{j2}}y'_j(x_2)\right), \qquad D_j = \frac{d}{dx_1}f_{y_{j1}} - \Omega_{y_j}(x_1),\\
E_j &= \frac{d}{dx_1}f_{y_{j2}}, \qquad F_j = \frac{d}{dx_2}f_{y_{j1}}, \qquad G_j = \frac{d}{dx_2}f_{y_{j2}} + \Omega_{y_j}(x_2),\\
H_{ij} &= f_{y_{i1}y_{j1}}, \qquad I_{ij} = f_{y_{i1}y_{j2}}, \qquad J_{ij} = f_{y_{i2}y_{j2}},\\
&\qquad (i, j = 1, \cdots, n).
\end{aligned}$$

If $y_i = y_i(x)$ is a minimizing arc it is evidently necessary that the first variation I_1,

$$I_1 \equiv \left(\frac{df_1}{d\epsilon}\right)\Bigg|^{\epsilon=0} = F_1(\xi, \eta),$$

vanish for all sets ξ, η satisfying the differential equations and end conditions (7) and (8). It is also necessary that the second variation I_2 be greater than or equal to zero for the same sets ξ, η.

3. The minimum problem of the second variation. A problem suggesting itself at this point is that of minimizing the second variation (14) in the class of all sets of variations ξ, η of class C''' on x_1x_2, satisfying the equations

$$(16) \qquad \begin{aligned}
&\frac{d\xi_\iota}{dx} = 0, \qquad \Phi_\alpha(x, \eta, \eta') = 0, \qquad && (\iota = 1, 2;\ \alpha = 1, \cdots, m),\\
&F_\sigma(\xi, \eta) = 0, && (\sigma = 2, \cdots, r).
\end{aligned}$$

It is evident that the second variation I_2 must be greater than or equal to zero in the class of all such sets ξ, η. It is found convenient to put a further restriction on the sets ξ, η. Let us introduce the equation

$$(17) \qquad \xi_1^2 + \xi_2^2 + \int_{x_1}^{x_2} \eta_i(x)\eta_i(x)\,dx = 1.$$

We then consider a second problem of the second variation and its relation to

the one first proposed. This second problem is to minimize the second variation in the class of all sets ξ, η satisfying the equations (16) and (17). We shall call ξ, η an admissible set for the first problem if the ξ's and η's are of class C''' on x_1x_2, and if the set satisfies the equations (16); and an admissible set for the second problem if the set further satisfies the equation (17). Every admissible set for the first problem, except the trivial one $\xi = \eta = 0$, will evidently give rise to a set $k\xi, k\eta$, which is an admissible set for the second problem, where k is a constant different from zero. Moreover, every admissible set ξ, η of the second problem is necessarily an admissible set of the first problem. It follows that the admissible sets of the second problem form a subclass of the class of all admissible sets of the first problem. We can then state the following

THEOREM. *If the second variation $I_2(\xi, \eta)$ is greater than or equal to zero for all sets ξ, η which satisfy the equations* (16), *where the ξ's and η's are of class C''' on x_1x_2, then $I_2(\xi, \eta)$ is necessarily greater than or equal to zero for all such sets ξ, η which satisfy both the equations* (16) *and* (17) *and conversely.*

The problem that will now be considered is that of minimizing the second variation $I_2(\xi, \eta)$ in the class of all sets ξ, η which satisfy the equations (16) and (17) with ξ's and η's of class C''' on x_1x_2. This problem is not quite in the form of the original problem in xy-space, but may be changed to that form by the introduction of new variables. For, let $\eta_0(x)$, $\eta_{n+1}(x)$ be defined by

$$\eta_0(x) = \int_{x_1}^{x} 2\omega(x, \eta, \eta')\,dx, \quad \eta_{n+1}(x) = \int_{x_1}^{x} \eta_i(x)\eta_i(x)\,dx,$$

$$\eta_0(x_1) = 0,\ \eta_0(x_2) = \int_{x_1}^{x_2} 2\omega(x, \eta, \eta')\,dx,\ \eta_{n+1}(x_1) = 0,\ \eta_{n+1}(x_2) = \int_{x_1}^{x_2} \eta_i\eta_i\,dx.$$

Let also the values of the constants ξ_1, ξ_2 at the end points x_1 and x_2 be denoted by ξ_{11}, ξ_{21} and ξ_{12}, ξ_{22} respectively, and consider (as we obviously may with no loss of generality) only ξ_{11} and ξ_{22} as occurring in the equations (15) and (16) and also in the second variation. We can then formulate an auxiliary Mayer problem as follows:

To find the properties of a set of functions

$$\eta_0(x),\ \eta_i(x),\ \eta_{n+1}(x),\ \xi_\iota(x), \qquad (i = 1, \cdots, \eta;\ \iota = 1, 2),$$

which minimizes the expression

$$g_1 \equiv I_2 = \eta_0(x_2) + Q(\xi_{11}, \xi_{22}, \eta_i(x_1), \eta_i(x_2)),$$

in the class of such sets of functions satisfying the differential equations

$$(18)\quad \begin{gathered} \eta'_0 - 2\omega(x,\eta,\eta') = 0, \qquad \Phi_\alpha(x,\eta,\eta') \equiv \phi_{\alpha y_j}\eta_j + \phi_{\alpha y'_j}\eta'_j = 0, \\ \eta'_{n+1} - \eta_i\eta_i = 0, \qquad \xi'_\iota = 0, \\ (i, j = 1, \cdots, n;\ \iota = 1, 2;\ \alpha = 1, \cdots, m), \end{gathered}$$

and the end conditions,

$$(19)\quad \begin{aligned} & g_2 \equiv \eta_0(x_1) = 0, \\ & g_{\sigma+1} \equiv F_\sigma(\xi_{11}, \xi_{22}, \eta_i(x_1), \eta_i(x_2)) = 0, \qquad (\sigma = 2, \cdots, r), \\ & g_{r+2} \equiv \eta_{n+1}(x_1) = 0, \\ & g_{r+3} \equiv \xi^2_{11} + \xi^2_{22} + \eta_{n+1}(x_2) - 1 = 0, \end{aligned}$$

with ξ's and η's of class C''' on x_1x_2.

This auxiliary problem is a Mayer problem of the type considered in the introduction and so a minimizing set of functions

$$\eta_0(x),\ \eta_i(x),\ \eta_{n+1}(x),\ \xi_\iota(x), \qquad (i = 1, \cdots, n;\ \iota = 1, 2),$$

of class C''' on x_1x_2 must satisfy the conditions there stated.[8] Let μ_0 be the multiplier associated with $\eta'_0 - 2\omega$; $\mu_\alpha(x)$, $(\alpha = 1, \cdots, m)$, those associated with Φ_α; μ_{m+1}, μ_{m+2}, those associated with ξ'_1, ξ'_2, respectively; and, finally, μ, that associated with $\eta'_{n+1} - \eta_i\eta_i$, where the μ's are of class C' on x_1x_2. Define Γ by the equation

$$2\Gamma \equiv \mu_0(\eta'_0 - 2\omega) + 2\mu_\alpha\Phi_\alpha + 2\mu_{m+1}\xi'_1 + 2\mu_{m+2}\xi'_2 + \mu(\eta'_{n+1} - \eta_i\eta_i).$$

Then the Euler differential equations for the auxiliary problem are

$$\frac{d}{dx}\Gamma_{\eta'_0} = \Gamma_{\eta_0}, \frac{d}{dx}\Gamma_{\eta'_i} = \Gamma_{\eta_i}, \frac{d}{dx}\Gamma_{\eta'_{n+1}} = \Gamma_{\eta_{n+1}}, \frac{d}{dx}\Gamma_{\xi'_1} = \Gamma_{\xi_1}, \frac{d}{dx}\Gamma_{\xi'_2} = \Gamma_{\xi_2}, \\ (i = 1, \cdots, n).$$

It follows from the $(n+2)$-nd of these equations that μ is a constant.

It should now be observed that a minimizing arc for this problem is surely a "normal" arc. For, from the form of the equations (18), it follows that the values of $\eta_0, \eta_{n+1}, \xi_1$ at $x = x_1$, and the value of ξ_2 at $x = x_2$, are entirely arbitrary. By hypothesis there exist r sets of admissible variations (ξ^ρ, η^ρ), $(\rho = 1, \cdots, r)$, so that the matrix $\| F_\sigma(\xi^\rho, \eta^\rho) \|$, $(\sigma = 2, \cdots, r)$ has rank $r - 1$. Hence from this and the arbitrariness of the end values of $\eta_0, \eta_{n+1}, \xi_1, \xi_2$, as just described, there must exist $r + 3$ sets of admissible variations $\eta_0, \eta_i, \eta_{n+1}, \xi_1, \xi_2$, such that the matrix

$$\| G_\nu(\xi^\delta, \eta^\delta) \|, \qquad (\nu = 2, \cdots, r+3;\ \delta = 1, \cdots, r+3),$$

has rank $r + 2$, where $G_\nu(\xi, \eta) = 0$ are the equations of variation corresponding to the equation (8).

---

[8] l.

Let $\bar{L}_\delta$ be the $r+3$ constants associated with g_δ with $\bar{L}_1 = 1$. Let θ take on successively the values $\eta_0, \eta_i, \eta_{n+1}, \xi_1, \xi_2$. Then from equations (4), the boundary conditions for the auxiliary minimum problem are seen to be

$$\bar{L}_\delta g_{\delta\theta_1} - 2\Gamma_{\theta'}(x_1) = 0, \qquad \bar{L}_\delta g_{\delta\theta_2} + 2\Gamma_{\theta'}(x_2) = 0.$$

From the three equations in which $\theta_2 = \eta_{n+1,2}$, $\theta_2 = \eta_{02}$, $\theta_1 = \eta_{n+1,1}$, it follows that $\bar{L}_{r+3} = -\mu$, $\mu_0 = -1$, $\bar{L}_{r+2} = \mu$, respectively. It then follows from the equation in which $\theta_1 = \eta_{01}$ that $\bar{L}_2 = -1$. When $\theta_2 = \xi_{12}$ and $\theta_1 = \xi_{21}$, it is seen that $\mu_{m+1} = \mu_{m+2} = 0$. Now re-name $\bar{L}_3, \cdots, \bar{L}_{r+1}$, respectively, $L_2, \cdots, L_r$. The remaining boundary conditions have the form

$$\begin{aligned} &Q_{\xi_{11}} + L_\sigma(f_{\sigma x_1} + f_{\sigma y_{i1}} y'_{i1}) - 2\mu\xi_{11} = 0,\\ &Q_{\eta_{i1}} + L_\sigma f_{\sigma y_{i1}} \qquad\qquad\quad - 2\Gamma_{\eta'_i}(x_1) = 0,\\ &Q_{\xi_{22}} + L_\sigma(f_{\sigma x_2} + f_{\sigma y_{i2}} y'_{i2}) - 2\mu\xi_{22} = 0,\\ &Q_{\eta_{i2}} + L_\sigma f_{\sigma y_{i2}} \qquad\qquad\quad + 2\Gamma_{\eta'_i}(x_2) = 0,\\ &\qquad (\sigma = 2, \cdots, r;\ i = 1, \cdots, n). \end{aligned} \tag{20}$$

The Euler differential equations of the auxiliary problem are now seen to be

$$\frac{d}{dx}\Gamma_{\eta'_i} = \Gamma_{\eta_i}, \tag{21}$$

where

$$\Gamma_{\eta'_i} = \omega_{\eta'_i} + \mu_\alpha \phi_{\alpha y'_i}, \qquad \Gamma_{\eta_i} = \omega_{\eta_i} + \mu_\alpha \phi_{\alpha y_i} - \mu\eta_i. \tag{22}$$

It is to be noted that the undetermined constants and multipliers are now $L_\sigma, \mu, \mu_\alpha(x)$ where $(\sigma = 2, \cdots, r;\ \alpha = 1, \cdots, m)$.

The results of this section may be summarized as follows:

Let $\xi_1, \xi_2, \eta_i(x)$ be a set of functions of class C''' on $x_1 \leqq x \leqq x_2$ which minimizes the second variation $I_2(\xi, \eta)$ in the class of such sets satisfying the differential equations and the end conditions (16) *and* (17). *Then there must exist m multipliers $\mu_\alpha(x)$, $(\alpha = 1, \cdots, m)$, of class C' and not all identically zero on x_1x_2, and r constants μ, L_σ, $(\sigma = 2, \cdots, r)$, not all zero, satisfying the differential equations* (21) *and the boundary conditions* (20), *in which ξ_{11} and ξ_{22} may be replaced by ξ_1 and ξ_2 respectively.*

4. The boundary value problem of the second variation and a necessary condition for the original problem. From the results of the last section it may be seen that there is a boundary value problem associated with the second variation which may be stated as follows:

To determine multipliers $\mu_\alpha(x)$ of class C' and constants L_σ, μ, together with variations $\xi_1, \xi_2, \eta_i(x)$ of class C''' on x_1x_2, which satisfy the differential equations

$$\frac{d}{dx}\Gamma_{\eta'_i} = \Gamma_{\eta_i}, \qquad \Phi_\alpha(x, \eta, \eta') = 0, \frac{d\xi_\iota}{dx} = 0, \tag{23}$$
$$(i = 1, \cdots, n;\ \alpha = 1, \cdots, m;\ \iota = 1, 2),$$

and the boundary conditions (20) and (8).

It will now be shown that there are restrictions upon the possible values of μ for which the boundary value problem has solutions, in view of the minimizing properties of the arc E for the original minimum problem considered. Suppose that the set ξ_ι, $\eta_i(x)$, L_σ, μ is a non-trivial solution of the boundary value problem and that it also satisfies the equation (17); this last condition could always be met for any given set by multiplying by a suitably chosen positive constant. Multiply the equations (20) by ξ_{11}, $\eta_i(x_1)$, ξ_{22}, $\eta_i(x_2)$, respectively, and add. Then since Q is a quadratic form in ξ_{11}, ξ_{22}, $\eta_i(x_1)$ and $\eta_i(x_2)$, and because of the equations (8), there will result

$$2Q + 2\eta_i\Gamma_{\eta'_i}\Big|_{x_1}^{x_2} - 2\mu(\xi_{11}^2 + \xi_{22}^2) = 0. \tag{24}$$

But from equations (22), (15) and (7), it follows that

$$\eta_i\Gamma_{\eta_i} + \eta'_i\Gamma_{\eta'_i} = \eta_i(\Gamma_{\eta_i} - \frac{d}{dx}\Gamma_{\eta'_i}) + \frac{d}{dx}(\eta_i\Gamma_{\eta'_i})$$
$$= 2\omega + \mu_\alpha\Phi_\alpha - \mu\eta_i\eta_i,$$

whence, in view of the equations (21),

$$\eta_i\Gamma_{\eta'_i}\Big|_{x_1}^{x_2} = \int_{x_1}^{x_2} 2\omega dx - \mu\int_{x_1}^{x_2}\eta_i\eta_i dx.$$

After substituting the right-hand side of this equation in the second term of (24) and recalling that

$$I_2 = \int_{x_1}^{x_2} 2\omega dx + Q,$$

it is seen that

$$2I_2 - 2\mu(\xi_{11}^2 + \xi_{22}^2 + \int_{x_1}^{x_2}\eta_i\eta_i dx) = 0,$$

and hence on account of (17) that

$$I_2 = \mu.$$

The following necessary condition on E has thus been proved:

If E is a minimizing arc for the original minimum problem in xy-space, then the boundary value problem of the second variation described in this section, the equations of which are (23), (20), *and* (8), *can have no solutions for negative values of the parameter μ.*

5. Transformation of the boundary value problem of the last section. The boundary value problem of the last section will now be transformed into a problem that has been discussed by Bliss.[9] It will first be assumed that a solution

$$\xi_\iota,\ \eta_i(x),\ \mu_\alpha(x),\ L_\sigma,\ \mu, \qquad (\iota=1,2;\ \alpha=1,\cdots,m;\ i=1,\cdots,n)$$

of the boundary value problem of the last section for a minimizing arc E for the original minimum problem has been found. Note that

$$F_{\sigma\xi_{11}}=f_{\sigma x_1}+f_{\sigma y_{i1}}y'_{i1},\quad F_{\sigma\xi_{22}}=f_{\sigma x_2}+f_{\sigma y_{i2}}y'_{i2},$$
$$F_{\sigma\eta_{i1}}=f_{\sigma y_{i1}},\quad F_{\sigma\eta_{i2}}=f_{\sigma y_{i2}}, \qquad (\sigma=2,\cdots,r).$$

For convenience, let ξ_1, ξ_2 be re-named η_{n+1}, η_{n+2}, respectively, and adjoin to the boundary value problem of the last section the equations

$$F_{r+1}\equiv\eta_{n+1}(x_1)-\eta_{n+1}(x_2)=0, \qquad F_{r+2}\equiv\eta_{n+2}(x_1)-\eta_{n+2}(x_2)=0.$$

It is then clear in view of these equations and hypothesis 4 of section 1 that the matrix

$$(25) \qquad \| F_{\tau\eta_{t1}}\ F_{\tau\eta_{t2}} \|, \qquad (\tau=2,\cdots,r+2;\ t=1,\cdots,n+2),$$

has rank $r+1$. As a matter of notation, let $\Gamma_{\eta'_i}$ be renamed ζ_i, and for convenience in subsequent proofs, let us introduce two new functions ζ_{n+1}, ζ_{n+2} defined by the equations

$$\zeta'_{n+1}=\zeta'_{n+2}=0, \qquad \zeta_{n+1}(x_1)=\zeta_{n+2}(x_1)=1.$$

The boundary conditions of an equivalent boundary value problem are then

$$(26)\qquad \begin{aligned}
&L_\sigma F_{\sigma\eta_{i1}} && +Q_{\eta_{i1}} && -2\zeta_{i1}=0,\\
&L_\sigma F_{\sigma\eta_{n+1,1}}+2F_{r+1,\eta_{n+1,1}} && +Q_{\eta_{n+1,1}}-2\mu\eta_{n+1,1} && -2\zeta_{n+1,1}=0,\\
&2F_{r+2,\eta_{n+2,1}} && +Q_{\eta_{n+2,1}} && -2\zeta_{n+2,1}=0,\\
&L_\sigma F_{\sigma\eta_{i2}} && +Q_{\eta_{i2}} && +2\zeta_{i2}=0,\\
&2F_{r+1,\eta_{n+1,2}} && +Q_{\eta_{n+1,2}} && +2\zeta_{n+1,2}=0,\\
&L_\sigma F_{\sigma\eta_{n+2,2}}+2F_{r+2,\eta_{n+2,2}} && +Q_{\eta_{n+2,2}}-2\mu\eta_{n+2,2} && +2\zeta_{n+2,2}=0,\\
&&& F_\tau=0,\\
&(\tau=2,\cdots,r+2;\ \sigma=2,\cdots,r;\ i=1,\cdots,n).
\end{aligned}$$

It is to be observed that F_σ and Q do not contain $\eta_{n+1,2}$ and $\eta_{n+2,1}$.

Suppose now that a_t and b_t, $(t=1,\cdots,n+2)$, are $2n+4$ constants satisfying the equations

[9] 3.

$$(27)\qquad a_t F_{\tau\eta_{t1}} + b_t F_{\tau\eta_{t2}} = 0, \qquad (\tau = 2, \cdots, r+2).$$

In view of the rank of the matrix (25), there will be

$$2n + 4 - (r+1) = 2n + 3 - r$$

linearly independent sets of such constants, say

$$a_{\lambda t},\ b_{\lambda t}, \qquad (\lambda = 1, 2, \cdots, 2n+3-r).$$

Multiply the first $n+2$ equations of (26) by $a_{\lambda t}$, the second set of $n+2$ equations by $b_{\lambda t}$, and add. The result will be because of (27)

$$\begin{gathered} a_{\lambda s} Q_{\eta_{s1}} - 2a_{\lambda s}\zeta_{s1} - 2a_{\lambda, n+1}\mu\eta_{n+1,1} \\ + b_{\lambda s} Q_{\eta_{s2}} + 2b_{\lambda s}\zeta_{s2} - 2b_{\lambda, n+2}\mu\eta_{n+2,2} = 0, \\ (\lambda = 1, \cdots, 2n+3-r;\ s = 1, \cdots, n+2). \end{gathered}$$

The quadratic form Q with arguments η_{s1}, η_{s2} may be written

$$\begin{gathered} Q(\eta(x_1), \eta(x_2)) = A_{st}\eta_{s1}\eta_{t1} + 2B_{st}\eta_{s1}\eta_{t2} + C_{st}\eta_{s2}\eta_{t2}, \\ (s, t = 1, \cdots, n+2), \end{gathered}$$

where A_{st} and C_{st} may without loss of generality be taken symmetric. It is to be emphasized that Q does not contain $\eta_{n+1,2}$, $\eta_{n+2,1}$. Now define $\| \delta'_{st} \|$ to be the matrix such that $\delta'_{st} = 1$ if $s = t = n+1$ and otherwise zero, and $\| \delta''_{st} \|$ to be the matrix such that $\delta''_{st} = 1$ if $s = t = n+2$ and otherwise zero. The boundary conditions (26) then imply the $2n+4$ conditions

$$\begin{gathered} -a_{\lambda s}\zeta_{s1} + b_{\lambda s}\zeta_{s2} + [a_{\lambda s}(A_{st} - \mu\delta'_{st}) + b_{\lambda s}B_{ts}]\eta_{t1} \\ + [b_{\lambda s}(C_{st} - \mu\delta''_{st}) + a_{\lambda s}B_{st}]\eta_{t2} = 0, \\ F_\tau(\eta) \equiv F_{\tau\eta_{s1}}\eta_{s1} + F_{\tau\eta_{s2}}\eta_{s2} = 0, \\ (\lambda = 1, \cdots, 2n+3-r;\ s, t = 1, \cdots, n+2;\ \tau = 2, \cdots, r+2). \end{gathered}$$

In matrix notation these equations may be written

$$(28)\qquad \begin{gathered} \left\| \begin{matrix} a_{\lambda s}(A_{st} - \mu\delta'_{st}) + b_{\lambda s}B_{ts}, & -a_{\lambda v} \\ F_{\tau\eta_{t1}}, & 0 \end{matrix} \right\| (\eta_{t1}, \zeta_{v1}) \\ + \left\| \begin{matrix} b_{\lambda s}(C_{st} - \mu\delta''_{st}) + a_{\lambda s}B_{st}, & b_{\lambda v} \\ F_{\tau\eta_{t2}}, & 0 \end{matrix} \right\| (\eta_{t2}, \zeta_{v2}) = 0, \\ (s, t, v = 1, \cdots, n+2;\ \lambda = 1, \cdots, 2n+3-r;\ \tau = 2, \cdots, r+2). \end{gathered}$$

The differential equations

$$(29)\qquad \begin{gathered} \frac{d}{dx}\Gamma_{\eta'_i} = \Gamma_{\eta_i}, \qquad \zeta'_{n+1} = \zeta'_{n+2} = 0, \\ \Phi_\alpha(x, \eta, \eta') \equiv \phi_{\alpha y_j}\eta_j + \phi_{\alpha y'_j}\eta'_j = 0, \qquad \eta'_{n+1} = \eta'_{n+2} = 0, \\ (i, j = 1, \cdots, n;\ \alpha = 1, \cdots, m), \end{gathered}$$

will now be reduced to an equivalent set which is for our purposes more useful.[10] Recalling (22) and (15), let H_{aj}, K_{aj} be defined with ζ_i by the equations

$$\zeta_i \equiv \Gamma_{\eta'_i}, \qquad H_{aj} \equiv \phi_{ay_j}, \qquad K_{aj} \equiv \phi_{ay'_j},$$

and assume as usual for regular problems in the calculus of variations that the determinant

$$D = \begin{vmatrix} R_{ij} & K_{\beta i} \\ K_{aj} & L_{a\beta} \end{vmatrix}, \qquad (\alpha, \beta = 1, \cdots, m;\ L_{a\beta} = 0),$$

is different from zero on $x_1 \leqq x \leqq x_2$. Then the $m + n$ equations

$$\zeta_i = Q_{ji}\eta_j + R_{ij}\eta'_j + K_{\beta i}\mu_\beta,$$
$$0 = H_{aj}\eta_j + K_{aj}\eta'_j,$$

can be solved for η'_j, μ_β. These solutions are

$$\eta'_i = -\frac{1}{D}(R_k{}^iQ_{jk}\eta_j + K_a{}^iH_{aj}\eta_j) + \frac{1}{D}R_j{}^i\zeta_j,$$
$$\mu_a = -\frac{1}{D}(K_a{}^kQ_{jk}\eta_j + L_\beta{}^aH_{\beta j}\eta_j) + \frac{1}{D}K_a{}^j\zeta_j,$$
$$(k = 1, \cdots, n),$$

in which, with reference to D, $R_k{}^i$ is the cofactor of R_{ki}; $K_a{}^i$, $L_\beta{}^a$, the cofactors, respectively, of K_{ai}, $L_{\beta a}$. Note that the matrix $\| R_k{}^i \|$ is symmetric. With the help of (29) it is seen that

$$\zeta'_i = P_{ij}\eta_j + Q_{ij}\eta'_j + H_{ai}\mu_a - \mu\eta_i.$$

When the values of η'_j and μ_a given above are substituted in these equations, they become

$$\zeta'_i = P_{ij}\eta_j - \frac{1}{D}Q_{ih}(R_k{}^hQ_{jk} + K_\nu{}^hH_{\nu j})\eta_j + \frac{1}{D}Q_{ih}R_j{}^h\zeta_j$$
$$-\frac{1}{D}H_{\nu i}(K_\nu{}^kQ_{jk} + L_\beta{}^\nu H_{\beta j})\eta_j + \frac{1}{D}H_{\nu i}K_\nu{}^j\zeta_j - \mu\eta_i,$$
$$(h = 1, \cdots, n;\ \nu = 1, \cdots, m).$$

If D is different from zero, the differential equations (29) are then equivalent to the following set in matrix notation:

---

[10] 15, p. 590.

$$
(30)\qquad \left\| \begin{matrix} P_{ql} - D^{-1}Q_{qs}(R_u{}^sQ_{lu} + K_\nu{}^sH_{\nu l}) - D^{-1}H_{\nu q}(K_\nu{}^uQ_{lu} + L_\beta{}^\nu H_{\beta l}), & -D^{-1}(R_s{}^qQ_{ls} + K_\alpha{}^qH_{\alpha l}), \\ D^{-1}(Q_{qs}R_l{}^s + H_{\nu q}K_\nu{}^l & D^{-1}R_l{}^q \end{matrix} \right\| (\eta_l, \zeta_l)
$$

$$
+ \mu \left\| \begin{matrix} 0 & 0 \\ -\bar{\delta}_{ql} & 0 \end{matrix} \right\| (\eta_l, \zeta_l) = \frac{d}{dx}(\eta_q, \zeta_q),
$$

$$
(l, q, s, u = 1, \cdots, n+2; \quad \alpha, \beta, \nu = 1, \cdots, m),
$$

where $\bar{\delta}_{ql}$ is the Kronecker symbol if q and l are $\leqq n$ and otherwise zero. If any one of l, q, s, u in a coefficient in (30) is greater than n, that coefficient is zero.

If D is different from zero, the boundary value problem of the last section can then be transformed into the following boundary value problem:

To determine functions $\eta_q = \eta_q(x)$, $\zeta_q = \zeta_q(x)$, $(q = 1, \cdots, n+2)$, with η's of class C''', ζ's of class C'' on x_1x_2, and a constant μ satisfying the differential equations (30) and the boundary conditions (28). It is clear that, under the hypotheses, every solution of the boundary value problem of section 4 whose equations are (23), (20), and (8), for a minimizing arc E of the original problem, is a solution of the transformed problem just stated.

6. The self-adjoint character of the auxiliary boundary value problem. The auxiliary boundary value problem of the last section whose equations are (30) and (28) will now be shown to be "self-adjoint" according to the definition of Bliss.[11] In the paper referred to, it is proved that a necessary and sufficient condition for the self-adjointness of a boundary value problem

$$
(31)\qquad \begin{gathered} \frac{dy_i}{dx} = (A_{i\alpha}(x) + \mu B_{i\alpha}(x))y_\alpha(x), \qquad x_1 \leqq x \leqq x_2, \\ M_{i\alpha}y_\alpha(x_1) + N_{i\alpha}y_\alpha(x_2) = 0, \end{gathered}
$$

is that there exist a transformation $T_{ik}(x)$ such that

$$
(32)\qquad \begin{gathered} T_{i\alpha}A_{\alpha k} + A_{\alpha i}T_{\alpha k} + T'_{ik} \equiv 0, \qquad T_{i\alpha}B_{\alpha k} + B_{\alpha i}T_{\alpha k} \equiv 0, \\ M_{i\alpha}T^{-1}_{\alpha\beta}(x_1)M_{k\beta} = N_{i\alpha}T^{-1}_{\alpha\beta}(x_2)N_{k\beta}, \\ (i, j, k, \alpha, \beta = 1, \cdots, n). \end{gathered}
$$

The functions $A_{ik}(x), B_{ik}(x), T_{ik}(x), T'_{ik}(x)$ are assumed to be continuous functions of x with $|T_{ik}| \neq 0$ on x_1x_2 and M_{ik}, N_{ik} are constants with n the rank of the matrix $\| M_{ik}, N_{ik} \|$.

The boundary value problem given by equations (28) and (30) is evidently of the form

[11] 3, p. 569.

$$
(33)\qquad
\begin{aligned}
&\frac{d}{dx}(y_i, z_i) = \left(\begin{Vmatrix} A_{ij} & B_{ij} \\ C_{ij} & D_{ij} \end{Vmatrix} + \mu \begin{Vmatrix} 0 & 0 \\ \bar{C}_{ij} & 0 \end{Vmatrix} \right) (y_j, z_j), \\
&\begin{Vmatrix} E_{pj} & F_{pj} \\ G_{lj} & 0 \end{Vmatrix} (y_j(x_1), z_j(x_1)) + \begin{Vmatrix} \bar{E}_{pj} & \bar{F}_{pj} \\ \bar{G}_{lj} & 0 \end{Vmatrix} (y_j(x_2), z_j(x_2)) = 0,
\end{aligned}
$$

where, in view of (28) and (27), we suppose

$$
(34)\qquad
\begin{gathered}
F_{p\alpha} G_{l\alpha} - \bar{F}_{p\alpha} \bar{G}_{l\alpha} = 0, \\
(i, j, \alpha = 1, \cdots, n;\ p = 1, \cdots, 2n - r;\ l = 1, \cdots, r).
\end{gathered}
$$

This is clearly of the same type as (31).

It will now be shown that with the transformation T, which with its inverse has the form

$$
T = \begin{Vmatrix} 0 & \delta_{ij} \\ -\delta_{ij} & 0 \end{Vmatrix}, \qquad T^{-1} = \begin{Vmatrix} 0 & -\delta_{ij} \\ \delta_{ij} & 0 \end{Vmatrix},
$$

the equations (32) are satisfied for the boundary value problem of the last section. To do this, we consider what properties the matrices $\| A_{ij} \|, \cdots, \| \bar{G}_{ij} \|$ of (33) must have if the equations (32) are to be satisfied. For the first set of equations of (32) to be satisfied, it is necessary and sufficient that

$$
\begin{Vmatrix} 0 & \delta_{i\alpha} \\ -\delta_{i\alpha} & 0 \end{Vmatrix} \cdot \begin{Vmatrix} A_{\alpha j} & B_{\alpha j} \\ C_{\alpha j} & D_{\alpha j} \end{Vmatrix} + \begin{Vmatrix} A_{\alpha i} & C_{\alpha i} \\ B_{\alpha i} & D_{\alpha i} \end{Vmatrix} \cdot \begin{Vmatrix} 0 & \delta_{\alpha j} \\ -\delta_{\alpha j} & 0 \end{Vmatrix} = 0,
$$

or

$$
\begin{Vmatrix} C_{ij} & D_{ij} \\ -A_{ij} & -B_{ij} \end{Vmatrix} + \begin{Vmatrix} -C_{ji} & A_{ji} \\ -D_{ji} & B_{ji} \end{Vmatrix} = 0.
$$

This is possible if and only if

$$
C_{ij} = C_{ji}, \qquad B_{ij} = B_{ji}, \qquad A_{ij} = -D_{ji};
$$

or, in other words, the matrices $\| C_{ij} \|$ and $\| B_{ij} \|$ must be symmetric, and the matrix $\| A_{ij} \|$ must be equal to the negative of the transpose of the matrix $\| D_{ij} \|$. It readily follows that the second set of equations (32) will be satisfied if and only if

$$
\bar{C}_{ij} = \bar{C}_{ji},
$$

that is, the matrix $\| C_{ij} \|$ must be symmetric. These conditions are easily verified to be fulfilled for the corresponding matrices of the equations (30). Hence for the transformation T, as defined above, the first two sets of equations (32) are satisfied for the boundary value problem of the last section.

The left member of the third equation of (32) for the problem (33) is

$$\begin{Vmatrix} E_{p\alpha} & F_{p\alpha} \\ G_{l\alpha} & 0 \end{Vmatrix} \cdot \begin{Vmatrix} 0 & -\delta_{\alpha\beta} \\ \delta_{\alpha\beta} & 0 \end{Vmatrix} \cdot \begin{Vmatrix} E_{q\beta} & G_{m\beta} \\ F_{q\beta} & 0 \end{Vmatrix}$$
$$= \begin{Vmatrix} F_{p\beta} & -E_{p\beta} \\ 0 & -G_{l\beta} \end{Vmatrix} \cdot \begin{Vmatrix} E_{q\beta} & G_{m\beta} \\ F_{q\beta} & 0 \end{Vmatrix},$$
$$(p, q = 1, \cdots, 2n - r; \quad l, m = 1, \cdots, r; \quad \alpha, \beta = 1, \cdots, n),$$

and this last product is equal to

$$\begin{Vmatrix} F_{p\beta}E_{q\beta} - E_{p\beta}F_{q\beta}, & F_{p\beta}G_{m\beta} \\ -G_{l\beta}F_{q\beta} & 0 \end{Vmatrix}.$$

Similarly the right member of the third equation of (32) for the problem (33) is found to be

$$\begin{Vmatrix} \bar{F}_{p\beta}\bar{E}_{q\beta} - \bar{E}_{p\beta}\bar{F}_{q\beta}, & \bar{F}_{p\beta}\bar{\bar{G}}_{m\beta} \\ -G_{l\beta}\bar{F}_{q\beta}, & 0 \end{Vmatrix}.$$

These two matrices will be equal if and only if

$$(35) \qquad \begin{gathered} F_{p\beta}G_{m\beta} = \bar{F}_{p\beta}\bar{\bar{G}}_{m\beta}, \\ F_{p\beta}E_{q\beta} - E_{p\beta}F_{q\beta} = \bar{F}_{p\beta}\bar{E}_{q\beta} - \bar{E}_{p\beta}\bar{F}_{q\beta}. \end{gathered}$$

The first equation is true because of (34). The second equation may be verified for the boundary value problem of the last section. For, let the range of the subscripts of the last equations be

$$\alpha, \beta = 1, \cdots, n + 2; \quad p, q = 1, \cdots, 2n + 3 - r; \quad l, m = 2, \cdots, r + 2.$$

Then substituting in (35) the values which make the matrices in (33) identical with those in (28), we find that the left and right members, respectively, of (35) are

$$(-a_{p\beta})(a_{qs}(A_{s\beta} - \mu\delta'_{s\beta}) + b_{qs}B_{\beta s}) - (a_{ps}(A_{s\beta} - \mu\delta'_{s\beta}) + b_{ps}B_{\beta s})(-a_{q\beta}),$$
$$(b_{p\beta})(b_{qs}(C_{s\beta} - \mu\delta''_{s\beta}) + a_{qs}B_{s\beta}) - (b_{ps}(C_{s\beta} - \mu\delta''_{s\beta}) + a_{ps}B_{s\beta})(b_{q\beta}).$$

Because of the symmetry of the matrices $\| A_{st} \|$, $\| C_{st} \|$, $\| \delta'_{st} \|$ and $\| \delta''_{st} \|$, these two expressions are equal if

$$a_{p\beta}b_{qs}B_{\beta s} = a_{ps}B_{s\beta}b_{q\beta},$$

and

$$b_{ps}B_{\beta s}a_{q\beta} = b_{p\beta}a_{qs}B_{s\beta}.$$

These, however, are true equations, as we may see by interchanging certain of the summation subscripts, and hence the third equation of (32) is satisfied for the boundary value problem of the last section.

All three of the equations (32) are then satisfied for the auxiliary boundary value problem of the last section, from which it follows that this problem is self-adjoint.

The auxiliary boundary value problem is, moreover, "definitely" self-adjoint, according to the definition of Bliss. Let S_{ik} be defined by $S_{ik}(x) = T_{\alpha i}B_{\alpha k}$. The boundary value problem (31) is then said to be "definitely" self-adjoint[12] if the following conditions are fulfilled: (1) The equations (32) are satisfied for the transformation $T_{ik}(x)$. (2) The matrix $\| S_{ik}(x) \|$ is symmetric. (3) The bilinear form $S_{\alpha\beta}(x)f_\alpha\bar{f}_\beta$ formed for a set of numbers f_i and their conjugate imaginaries $\bar{f}_i$ is positive or zero at every point of x_1x_2. (4) This form vanishes identically for a set of solutions $f_i(x)$ of a system of equations

$$f'_i(x) = A_{i\alpha}(x)f_\alpha(x) + B_{i\alpha}(x)g_\alpha(x), \qquad (i, k, \alpha, \beta = 1, \cdots, n),$$

where the g_i's are continuous functions of x on x_1x_2 but otherwise arbitrary, only when the functions $f_i(x)$ are all identically zero. These conditions are easily seen to be satisfied for the auxiliary boundary value problem of the last section since S has the value

$$S = \begin{Vmatrix} 0 & -\delta_{qv} \\ \delta_{qv} & 0 \end{Vmatrix} \cdot \begin{Vmatrix} 0 & 0 \\ -\bar{\delta}_{vl} & 0 \end{Vmatrix} = \begin{Vmatrix} \bar{\delta}_{ql} & 0 \\ 0 & 0 \end{Vmatrix},$$
$$(q, v, l = 1, \cdots, n+2).$$

The theory of definitely self-adjoint boundary value problems as developed by Bliss[13] may now be applied in full to the problem of the last section. According to that theory, the characteristic constants μ for which this problem has solutions must all be real and are denumerably infinite in number, and the linearly independent characteristic solutions corresponding to each characteristic constant may be chosen real. Since every solution of the problem of section 4 is also a solution of the problem of the last section, it follows that the characteristic constants μ for that problem are also all real, and the linearly independent characteristic solutions corresponding to each characteristic constant may be chosen real. By the criterion of section 4, none of the characteristic parameter values for the problem of section 4 can be negative if E is a minimizing arc for the original problem in xy-space.

MARIETTA COLLEGE,
MARIETTA, OHIO.

[12] 3, p. 570.

[13] 3, p. 571.

BIBLIOGRAPHY.

1. Bliss, "The problem of Mayer with variable end points," *Transactions of the American Mathematical Society*, vol. 19 (1918), pp. 305-314.

2. Bliss, "A boundary value problem of the calculus of variations," *Bulletin of the American Mathematical Society*, vol. 32 (1926), pp. 317-331.

3. Bliss, "A boundary value problem for a system of ordinary linear differential equations of the first order," *Transactions of the American Mathematical Society*, vol. 28 (1926), pp. 561-584.

4. Cope, "An analogue of Jacobi's condition for the problem of Mayer with variable end points," *Dissertation* (1927), The University of Chicago.

5. Bliss, "The problem of Lagrange in the calculus of variations," *American Journal of Mathematics*, vol. 52 (1930), pp. 673-744.

6. Morse and Myers, "The problems of Lagrange and Mayer with variable end points," *Proceedings of the American Academy of Arts and Sciences*, vol. 66 (1931), pp. 235-253.

7. Morse, "Sufficient conditions in the problem of Lagrange with variable end conditions," *American Journal of Mathematics*, vol. 53 (1931), pp. 517-546.

8. Bliss, "The problem of Bolza in the calculus of variations," *Annals of Mathematics*, vol. 33 (1932), pp. 261-274.

9. Myers, "Adjoint systems in the problem of Mayer under general end-conditions," *Bulletin of the American Mathematical Society*, vol. 38 (1932), pp. 303-312.

10. Reid, "A boundary value problem associated with the calculus of variations," *American Journal of Mathematics*, vol. 54 (1932), pp. 769-780.

11. Bliss and Hestenes, "Sufficient conditions for a problem of Mayer in the calculus of variations," *Transactions of the American Mathematical Society*, Vol. 35 (1933), pp. 305-326.

12. Hestenes, "Sufficient conditions for a problem of Mayer with variable end points," *Transactions of the American Mathematical Society*, vol. 35 (1933), pp. 479-490.

13. Morse, "The calculus of variations in the large," *American Mathematical Society Colloquium Publications*, vol. 18 (1934).

14. Reid, "Analogues of the Jacobi condition for the problem of Mayer in the calculus of variations," *Annals of Mathematics*, vol. 35 (1934), pp. 836-848.

15. Bolza, *Vorlesungen über Variationsrechnung* (1909).

ON THE ASYMPTOTIC DISTRIBUTION OF $\zeta'/\zeta(s)$ IN THE CRITICAL STRIP.*

By Richard Kershner and Aurel Wintner.

According to Bohr,[1] Euler's product for $\zeta(s)$, which is divergent for $\sigma < 1$, has for $\sigma > \frac{1}{2}$ certain convergence tendencies. In particular, the formal expansion of $\log \zeta(s)$ is, for every fixed $\sigma > \frac{1}{2}$, convergent in relative measure to $\log \zeta(s)$, and this holds uniformly for all $\sigma >$ const. $> \frac{1}{2}$. Bohr's proof of this fact consists in first replacing $\zeta(s)$ by

$$(1 - 2^{1-s})\zeta(s) = \sum_{n=1}^{\infty} (-1)^{n+1}/n^s, \text{ where } \sigma > 0,$$

then applying Schnee's mean-value theorem uniformly for $\sigma >$ const. $> \frac{1}{2}$, and finally removing the factor $(1 - 2^{1-s})$. Thus the method does not apply to the function[2] $\zeta'/\zeta(s)$, a function more directly connected with the prime number distribution than $\zeta(s)$ itself.

In the present note there will first be obtained results for $\zeta'/\zeta(s)$ which are analogous to those of Bohr for $\log \zeta(s)$. This will be made possible in view of a general principle which may be formulated roughly as follows: One cannot lose uniform convergence in relative measure by term-by-term differentiation of a sequence of analytic functions (the same does not hold for term by term integration).

The result is then applied to obtain for the asymptotic distribution of $\zeta'/\zeta(s)$, where $\sigma > \frac{1}{2}$, results which are analogous to those previously obtained[3] for $\log \zeta(s)$ or $\zeta(s)$. Due to the convergence in relative measure, these results follow immediately from the general theory of infinite convolutions,[3] since the logarithms of the prime numbers are linearly independent. The asymptotic distribution of $\zeta'/\zeta(s)$ has recently[4] been discussed for $\sigma > 1$. The facts to be proved seem to be new not only for $\frac{1}{2} < \sigma < 1$ but for $\sigma = 1$ as well.

The results are independent of Riemann's hypothesis.

Let $s_n = \sigma_n + it_n$ denote those zeros, if any, of $\zeta(s)$ for which $\sigma_n > \frac{1}{2}$, and let J_n be the interval

$$t = t_n, \qquad \tfrac{1}{2} < \sigma \leqq \sigma_n \ (< 1),$$

* Received January 29, 1937.

[1] Bohr [1], [2].

[2] By $\zeta'/\zeta(s)$ is meant $\zeta'(s)/\zeta(s)$.

[3] Jessen and Wintner [3].

[4] van Kampen and Wintner [4], Section 6.

an interval which is perpendicular to the critical line. Let Γ denote the set obtained from the half-plane $\sigma > \frac{1}{2}$ by removing the segments J_n and also the segment

$$t = 0, \qquad \tfrac{1}{2} < \sigma \leqq 1.$$

Thus Γ is open, simply-connected, and contains the half-plane $\sigma > 1$. On placing

$$\zeta_n(s) = \zeta(s) \prod_{k=1}^{n} (1 - p_k^{-s}); \qquad (n = 0, 1, 2, \cdots) \tag{1}$$

it is clear that

$$\zeta_n(s) \neq 0 \text{ in } \Gamma \qquad (\zeta_0 = \zeta). \tag{2}$$

Let $\log \zeta_n(s)$ denote that logarithm of $\zeta_n(s)$ which is regular analytic in Γ and vanishes as $\sigma \to +\infty$. The considerations of the sequel will be based on a classical result concerning the manner of divergence of Euler's product in the critical strip, namely on

Bohr's Lemma.[5] Corresponding to three arbitrary numbers η, ϵ_1, ϵ_2 satisfying

$$0 < \eta < \tfrac{1}{2}; \quad 0 < \epsilon_1 < 1, \quad 0 < \epsilon_2 < 1,$$

there exists an $N = N(\eta, \epsilon_1, \epsilon_2)$ with the property that one can choose for every $n \geqq N$ a u_n such that if $T \geqq u_n$, then the interval $2 \leqq \tau \leqq T$ contains a finite number of mutually disjoint subintervals I which have a total length greater than $(1 - \epsilon_2)T$ and are such that if τ is in an I, then on the one hand the half-strip

$$\sigma \geqq \tfrac{1}{2} + \eta, \qquad \tau - \tfrac{1}{2} \leqq t \leqq \tau + \tfrac{1}{2}$$

of the $(\sigma + it)$-plane is contained in Γ and on the other hand

$$|\log \zeta_n(s)| \leqq \epsilon_1$$

for every $s = \sigma + it$ in any of these half-strips.

Let Σ_0 denote the boundary of a square of side $\frac{1}{2}$ in the complex s-plane and let Σ_δ, where $0 < \delta < \frac{1}{4}$, denote the boundary of that square of side $\frac{1}{2} - 2\delta$ which is symmetrically placed in Σ_0. Then, if $f(s)$ is any function regular analytic on and within Σ_0, its derivative $f'(s)$ satisfies the inequality

$$\max_{\Sigma_\delta} |f'(s)| \leqq \frac{1}{\delta^2} \max_{\Sigma_0} |f(s)|, \qquad (0 < \delta < \tfrac{1}{4}), \tag{3}$$

regardless of the position of Σ_0 in the s-plane. In fact, if s is in the interior of Σ_0, then, according to Cauchy,

[5] Bohr [1], Hilfssatz 5, p. 82. The numbers denoted above by $\epsilon_1, \epsilon_2, \eta, N, u_n$ are in Bohr's notation $\epsilon', \epsilon'', \sigma' - \frac{1}{2}, N^*, T^*$ respectively.

$$2\pi i f'(s) = \int_{\Sigma_0} (s-w)^{-2} f(w)\,dw.$$

Now if s is on or within Σ_δ, then $|s-w| \geqq \delta$ for every w on Σ_0, so that (3) is obvious. Let Σ_0 be any square of side $\frac{1}{2}$ which has one side on the line $\sigma = \frac{1}{2} + \eta$ and is contained in one of the half-strips mentioned in Bohr's Lemma. On applying (3) to each of these squares Σ_0 and to the function $f(s) = \log \zeta_n(s)$, it is seen that Bohr's Lemma implies the following:

Corresponding to four arbitrary numbers δ, η, ϵ_1, ϵ_2 satisfying

$$0 < \delta < \tfrac{1}{4}, \quad 0 < \eta < \tfrac{1}{2}, \quad 0 < \epsilon_1 < 1, \quad 0 < \epsilon_2 < 1,$$

there exists an $N = N(\eta, \epsilon_1, \epsilon_2)$ with the property that one can choose for every $n \geqq N$ a u_n such that if $T \geqq u_n$, then the interval $2 \leqq \tau \leqq T$ contains a finite number of mutually disjoint subintervals I which have a total length greater than $(1-\epsilon_2)T$ and are such that if τ is in an I, then on the one hand the square

$$\tfrac{1}{2} + \eta + \delta \leqq \sigma \leqq 1 + \eta - \delta, \qquad \tau - \tfrac{1}{4} + \delta \leqq t \leqq \tau + \tfrac{1}{4} - \delta$$

is contained in Γ and on the other hand

$$|\zeta'_n/\zeta_n(s)| \leqq \epsilon_1/\delta^2$$

for every $s = \sigma + it$ in any of these squares.

Now if σ_0 is a fixed value such that $\frac{1}{2} < \sigma_0 \leqq 1$, one can choose $\eta > 0$ and $\delta > 0$ such that

$$\tfrac{1}{2} + \eta + \delta < \sigma_0 < 1 + \eta - \delta, \qquad (\eta < \tfrac{1}{2},\ \delta < \tfrac{1}{4}).$$

Let δ and η be fixed. Then, on placing $\epsilon^1 = \epsilon_1/\delta^2$ and applying the above result only for a fixed $\sigma = \sigma_0$, one obtains the following corollary: If $\sigma_0 > \frac{1}{2}$ is fixed,[6] there exists for every pair of positive numbers ϵ^1, ϵ_2 an $N = N(\epsilon^1, \epsilon_2)$ with the property that one can choose for every $n \geqq N$ a u_n such that if $T \geqq u_n$, then

$$|\zeta'_n/\zeta_n(\sigma_0 + it)| \leqq \epsilon^1$$

whenever t is on a certain subset of the interval $2 \leqq t \leqq T$, and this subset has a measure greater than $(1-\epsilon_2)T$. Since ϵ^1, ϵ_2 are independent of each other, it follows, by letting $\epsilon_2 \to 0$, $T \to +\infty$ and writing σ and ϵ instead of σ_0 and ϵ^1, that, if $\sigma > \frac{1}{2}$ is fixed, then, for every fixed $\epsilon > 0$,

$$\lim_{n\to+\infty} \limsup_{T\to+\infty} \frac{1}{T} \operatorname*{meas}_{0<t<T} \{|\zeta'_n/\zeta_n(\sigma+it)| > \epsilon\} = 0, \tag{4}$$

where the factor of $1/T$ denotes the linear measure of the set of those values

[6] If $\sigma_0 > 1$, the statement is trivial.

t for which

$$0 < t < T \quad \text{and} \quad |\zeta'_n/\zeta_n(\sigma + it)| > \epsilon.$$

Thus (4) means that if $\sigma > \frac{1}{2}$ is fixed, the sequence

$$\zeta'_1/\zeta_1(\sigma + it), \cdots, \zeta'_n/\zeta_n(\sigma + it), \cdots$$

tends in relative measure [7] to the function of t which is $\equiv 0$. It is clear from the above proof that the convergence in relative measure is uniform for all $\sigma \geqq \bar{\sigma}$, if $\bar{\sigma} > \frac{1}{2}$ is fixed. On writing

$$f_n(t) [\rightarrow] f(t), \qquad n \rightarrow +\infty,$$

if $f_n(t)$ tends in relative measure to $f(t)$, i. e., if

$$\lim_{n\to+\infty} \limsup_{T\to+\infty} \frac{1}{T} \operatorname*{meas}_{0<t<T} \{|f(t) - f_n(t)| > \epsilon\} = 0$$

holds for every fixed $\epsilon > 0$, the above result may be formulated as follows:

THEOREM I. *If $\sigma > \frac{1}{2}$ is fixed, then Euler's series*

$$(5) \qquad -\sum_{k=1}^{\infty} \frac{p_k^{-s} \log p_k}{1 - p_k^{-s}}, \qquad (s = \sigma + it),$$

for the function $\zeta'/\zeta(s)$ converges in relative measure to $\zeta'/\zeta(\sigma + it)$, i. e.,

$$(6) \qquad \rho_n(\sigma + it) [\rightarrow] \zeta'/\zeta(\sigma + it), \qquad n \rightarrow +\infty,$$

where ρ_n is an abbreviation for

$$(7) \qquad \rho_n(s) = \rho_n(\sigma + it) = -\sum_{k=1}^{n} \frac{p_k^{-s} \log p_k}{1 - p_k^{-s}}.$$

Furthermore, (4) *holds uniformly for $\sigma \geqq \bar{\sigma}$, if $\bar{\sigma} > \frac{1}{2}$ is fixed.*

In fact, if $\sigma > 1$, then (5) may be written as the absolutely convergent Dirichlet series of $\zeta'/\zeta(s)$, and so

$$(8) \qquad \zeta'_n/\zeta_n(s) = \zeta'/\zeta(s) - \rho_n(s)$$

is, for $\sigma > 1$, obvious from the definitions (6) and (1) of $\rho_n(s)$ and $\zeta_n(s)$. Now the function (7) is regular analytic in the half-plane $\sigma > 0$, and this half-plane contains the domain Γ defined above. Since $\zeta(s)$ and $\zeta_n(s)$ are regular analytic and distinct from zero in the simply connected domain Γ which contains the half-plane $\sigma > 1$, it follows that (8) holds at every point s of Γ. Consequently, (6) is equivalent to (4).

---

[7] Cf. Jessen and Wintner [3], Section 11.

It may be mentioned that the function $\zeta'/\zeta(\sigma+it)$ to which the series (5) is convergent in relative measure is a continuous function of t, if $\sigma>\frac{1}{2}$ is not the abscissa of a zero of ζ (supposing that there exist such zeros). The corresponding fact does not hold in Bohr's case of $\log\zeta(s)$, if the Riemann hypothesis is false. In fact, if the Riemann hypothesis is false, then each point of the cuts J_n occurring in the definition of Γ is a discontinuity of the function $\log\zeta(\sigma+it)$ of t, if $\sigma>\frac{1}{2}$ is fixed and $\sigma<\sigma_n$, where σ_n is the abscissa of the right-hand end of J_n.

Remark. Since $\zeta'/\zeta(s)$ is regular for $s=1+it\neq 1$, and since the series

$$\sum_{n=1}^{\infty}\{(1-p_n^{-s})^{-1}p_n^{-s}\log p_n-p_n^{-s}\log p_n\}=\sum_{n=1}^{\infty}\sum_{k=2}^{\infty}p_n^{-ks}\log p_n$$

represents a regular function for $\sigma>\frac{1}{2}$ and so for $s=1+it$, it is clear from

$$\sum_{n=1}^{m}p_n^{-1}\log p_n\sim\log p_m=O(\log m),\qquad m\to+\infty,$$

that the series

$$-\sum_{n=1}^{\infty}p_n^{-s}\log p_n=\zeta'/\zeta(s)+\sum_{n=1}^{\infty}\sum_{k=2}^{\infty}p_n^{-ks}\log p_n$$

is, in virtue of a general theorem of M. Riesz,[8] convergent for every $s=1+it\neq 1$. In other words, the series (5) is convergent for every $s=1+it\neq 1$. This does not imply, however, the statement of Theorem I for $\sigma=1$, since (5) is not uniformly convergent for all $s=1+it$, $2<t<+\infty$ (if $\sigma<1$, then (5) is clearly divergent for every t). Correspondingly, the following theorem seems to be new not only for $\frac{1}{2}<\sigma\leq 1$ but for $\sigma=1$ as well:

THEOREM II. *The function $\zeta'/\zeta(\sigma+it)$ possesses, for every fixed $\sigma>\frac{1}{2}$, an asymptotic distribution function.*

This theorem is[9] a consequence[10] of Theorem I, since the function $\rho_n(\sigma+it)$ of t is, according to (7), almost periodic in the sense of Bohr.

THEOREM III. *The asymptotic distribution function of $\zeta'/\zeta(\sigma+it)$, where $\sigma>\frac{1}{2}$ is fixed, is absolutely continuous with a density which possesses partial derivatives of arbitrarily high order. If $\frac{1}{2}<\sigma<1$, then the density is a transcendental entire function of two variables.*

---

[8] M. Riesz [5], p. 350.

[9] Cf. Jessen and Wintner [3], Section 11.

[10] The statement of Theorem I as to uniformity with respect to σ is not needed.

In fact, since the logarithms of the prime numbers are linearly independent, Theorem III follows from Theorem I by an obvious modification of methods previously applied [11] to $\log \zeta(\sigma + it)$. The same holds [11] also for

THEOREM IV. *If* $\frac{1}{2} < \sigma \leqq 1$, *the density of the asymptotic distribution function of the function* $\zeta'/\zeta(\sigma + it)$ *of* t *is everywhere positive but vanishes in the infinity as strongly as a Gaussian density.*

Needless to say, Theorem IV may be considered as an indication of the truth of Riemann's hypothesis (which, in turn, does not imply Theorem IV).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] H. Bohr, "Zur Theorie der Riemannschen Zetafunktion im kritischen Streifen," *Acta Mathematica*, vol. 40 (1915), pp. 67-100.

[2] ———, "Über diophantische Approximationen und ihre Anwendungen auf Dirichletsche Reihen, besonders auf die Riemannsche Zetafunktion." *Proceedings of the Fifth Congress of Scandinavian Mathematicians*, Helsingfors, 1922, pp. 131-154.

[3] B. Jessen and A. Wintner, "Distribution functions and the Riemann zeta function," *Transactions of the American Mathematical Society*, vol. 38 (1935), pp. 48-88.

[4] E. R. van Kampen and A. Wintner, "Convolutions of distributions on convex curves and the Riemann zeta function," *American Journal of Mathematics*, vol. 59 (1937), pp. 175-204.

[5] M. Riesz, "Ein Konvergenzsatz für Dirichletsche Reihen," *Acta Mathematica*, vol. 40 (1915), pp. 349-361.

[11] Jessen and Wintner [3], Sections 8, 13, 14.

ON THE ADDITION OF CONVEX CURVES AND THE DENSITIES OF CERTAIN INFINITE CONVOLUTIONS.*

By E. R. van Kampen.

Let $S_1, S_2, \cdots$ be a sequence of convex curves in a plane such that the infinite convolution ϕ of the distribution functions [1] $\phi_1, \phi_2, \cdots$ corresponding to $S_1, S_2, \cdots$ is convergent. Under very general conditions it has been proved [2] that the distribution function ϕ has continuous partial derivatives of arbitrarily high order. If the curves $S_1, S_2, \cdots$ are analytic, the question arises whether ϕ is regular analytic anywhere in the plane. In certain special cases this question has already been investigated.[3] The object of the present paper is to develop a concise formalism for the description of the geometrical process of addition of convex curves. On using this formalism in connection with the study of ϕ a simple proof will be given in Section IV for the analyticity of the density of ϕ in certain specified regions. Although the boundaries of these regions are defined in such a way as to suggest their having a singular character for the density of ϕ, it still remains a problem to decide under fairly general conditions whether or not it is possible for ϕ to be regular on these boundaries.[4] As an application it will be shown in Section V that there exist for the asymptotic distribution function of the logarithm of the Riemann zeta function in addition to the ring shaped regions of analyticity previously determined,[5] certain regions of analyticity which may be described as crescent shaped.

In Sections I, II, III, in which the geometric considerations are given, the convex curves have been replaced by convex hypersurfaces in an n-dimensional space. This modification has no effect on the proofs but justifies the introduction of vector notations which lead to a slight simplification even if $n = 2$. The content of Section I is in the main a repetition of a treatment of the same problem by Kershner.[6] However, apart from the minor differences concerning the dimension of the containing space and the notation, the complete induction process used by Kershner has been replaced by a direct passage from the case

* Received March 15, 1937.

[1] Cf. Jessen and Wintner [3], Sections 4 and 7.

[2] *Ibid.*, Section 8.

[3] Van Kampen and Wintner [6].

[4] *Ibid.*, Appendix.

[5] *Ibid.*, Section 7.

[6] Kershner [4].

of 2 hypersurfaces to the general case, so that a limiting process for the case of infinitely many hypersurfaces becomes superfluous. The result of Section III is not used in what follows, but gives an insight in the character of the sets F and G introduced earlier.

I. Vectorial sums of convex hypersurfaces. Let ω, η be variable vectors in a real finite-dimensional vector space such that ω is of length 1 and let $\omega \cdot \eta$ denote the scalar product of these two vectors. By a convex hypersurface will be understood the point set theoretic boundary of a non-empty bounded convex point set in the space. Thus in particular a convex hypersurface may degenerate into a single point.

Let S_k be a convex hypersurface. If the exterior normal to a supporting hyperplane of S_k at the point η of S_k has the direction determined by ω, the notation

$$\eta = \eta_k(\omega) \tag{1}$$

will be used, even though (1) is in general not an allowable parametric representation of S_k. The supporting function $h_k(\omega)$ of S_k appears in the form

$$h_k(\omega) = \omega \cdot \eta_k(\omega). \tag{2}$$

The convex hypersurface S^*_k which is symmetrical to S_k with respect to the origin may be represented by

$$\eta = -\eta_k(-\omega), \tag{3}$$

and so its supporting function is

$$h^*_k(\omega) = h_k(-\omega). \tag{4}$$

The fact that S_k is convex may be expressed by the inequality

$$\omega \cdot \eta_k(\omega') \leqq h_k(\omega), \tag{5}$$

where the equality sign holds for $\omega = \omega'$. On substituting $-\omega$ in (5) instead of ω one obtains

$$-h_k(-\omega) \leqq \omega \cdot \eta_k(\omega') \tag{6}$$

where the equality sign holds for $\omega = -\omega'$, in view of (2). It follows from (5) and (6) that

$$h_k(\omega) + h_k(-\omega) \geqq 0, \tag{7}$$

which relation is also evident from the convexity of S_k. A point η clearly is contained in the closed convex set determined by S_k if and only if

(8) $$\omega \cdot \eta \leqq h_k(\omega),$$

for every ω, while such a point η is a point $\eta = \eta_k(\omega')$ of S_k if and only if the equality sign in (8) holds for $\omega = \omega'$.

Let an infinite sequence of convex hypersurfaces

(9) $$S_1, S_2, \cdots$$

be given by (1) such that

(10) $$|\eta_k(\omega)| < a_k, \qquad \Sigma_k a_k < +\infty,$$

for certain numbers a_k and all ω. Thus all series occurring in this Section are absolutely (uniformly) convergent. The particular case where only a finite number of hypersurfaces are given need not be excluded but will not be referred to. By a simple argument [7] it may be seen that the locus S represented by [8]

(11) $$\eta = \Sigma_k \eta_k(\omega)$$

is a convex hypersurface and has the supporting function

(12) $$h(\omega) = \Sigma_k h_k(\omega).$$

The locus represented by

(13) $$\eta = \Sigma_k \eta_k(\omega_k),$$

where the ω_k vary independently, is the vectorial sum of the S_k and will be denoted by T. It follows from (5) and (13) by summation that

(14) $$\omega \cdot \eta \leqq \Sigma_k h_k(\omega) = h(\omega),$$

for every point η in T, while in view of (2) the equality sign holds in (14) if $\omega_k = \omega$ for every k, in which case η is a point in S also. Thus T is contained in the closed convex set C determined by S. On adding (2) for $k = 1$ and (6) for all $k > 1$, it follows from (13) that

$$h_1(\omega) - \sum_{k>1} h_k(-\omega) \leqq \omega \cdot \eta$$

if η is determined by (13) and $\omega = \omega_1$. Thus the set U of points η defined by the inequality

(15) $$\omega \cdot \eta < h_1(\omega) - \sum_{k>1} h_k(-\omega)$$

[7] Haviland [2].

[8] This equation should be read in the sense that if for a given ω and for one or more values of k the functions $\eta_k(\omega)$ are not univalued, then (11) represents all possible values taken by the sum on the right.

does not have a point in common with T although U is contained in C as a consequence of (7). From the form of (15) it is clear that the set U is either empty or an open convex set. On replacing ω in (15) by $-\omega$ and adding the result to (15) one obtains

$$h_1(\omega) + h_1(-\omega) > \sum_{k>1} h_k(\omega) + h_k(-\omega), \tag{16}$$

so that (16) is a necessary (but not sufficient) condition for the set U to be non-empty. It follows also that U must be empty except for at most one choice of S_1 among the given S_k. In order to assure that the correct choice of S_1 has been made, it will be assumed that the S_k have been enumerated in such a way that

$$\operatorname{Max}_\omega\{h_k(\omega) + h_k(-\omega)\} \leqq \operatorname{Max}_\omega\{h_{k+1}(\omega) + h_{k+1}(-\omega)\}, \quad (k = 1, 2, \cdots) \tag{17}$$

A boundary point η_0 of U clearly satisfies the inequality

$$\omega \cdot \eta_0 \leqq h_1(\omega) - \sum_{k>1} h_k(-\omega), \tag{18}$$

while the equality sign in (18) holds for at least one value of ω, say for $\omega = \omega_0$. Put

$$\eta' = \sum_{k>1} \eta_k(-\omega_0) \quad \text{and} \quad \eta_0 = \eta' + \eta''. \tag{19}$$

On placing in (6) the variable ω' equal to $-\omega_0$ and adding the result to (18) for all $k > 1$, one obtains in view of (19)

$$\omega \cdot \eta_0 - \omega \cdot \eta' = \omega \cdot \eta'' \leqq h_1(\omega),$$

where the equality sign still holds for $\omega = \omega_0$. Thus $\eta' = \eta_1(\omega_0)$, by the remark made in connection with (8), so that η_0 is a point of the locus R' represented by

$$\eta = \eta_1(\omega) + \sum_{k>1} \eta_k(-\omega). \tag{20}$$

Since R' is contained in T it is clear that the boundary R of U is contained in T. Now it will be shown that $T = C - U$.

Consider first the case where $k = 1, 2$ only and let η_0 be a point which is in C but not in T. Obviously

$$\eta_0 - \eta_2(-\omega_2) \neq \eta_1(\omega_1), \tag{21}$$

for every ω_1, ω_2, so that the vectorial sum T' of the convex hypersurfaces represented by the point η_0 and by the convex hypersurface S^*_2 does not have a point in common with S_1. Obviously T' is obtained from S^*_2 by a trans-

lation along the vector η_0. It is impossible that T' and S_1 are exterior to each other, since otherwise the inequality (21) would hold for all η on a line joining η_0 with ∞, hence no point of this line would be in T. This contradicts the assumption that η_0 is in C, since the boundary of C is in T. Also S_1 cannot be interior to T', by (17). Thus T' must be interior to S_1. This implies

$$\omega \cdot \eta_0 + h_2(-\omega) < h_1(\omega), \tag{22}$$

so that η_0 is a point of U, and the statements that $T = C - U$ is proved if only two hypersurfaces are considered.

In the general case,[9] let again η_0 be a point in C but not in T, so that

$$\omega \cdot \eta_0 < \Sigma_k\, h_k(\omega), \tag{23}$$

for all ω, since η_0 is in C but not on the boundary S of C and

$$\eta_0 \neq \Sigma_k\, \eta_k(\omega_k), \tag{24}$$

for any sequence $\omega_1, \omega_2, \cdots$. It must be shown that η_0 satisfies (15). The inequality (24) may be written in the form

$$\eta_0 - \Sigma''\, \eta_k(-\omega_k) \neq \Sigma'(\omega_k), \tag{25}$$

where Σ' denotes a summation over a set of distinct positive integers k and Σ'' denotes the summation over the remaining positive integers k. It may be seen from (25) (by a repetition of the argument which follows (21)) that either

$$\omega \cdot \eta_0 > \Sigma'\, h_k(\omega) - \Sigma''\, h_k(-\omega) \tag{26a}$$

or

$$\omega \cdot \eta_0 < \Sigma'\, h_k(\omega) - \Sigma''\, h_k(-\omega) \tag{26b}$$

holds for given summations Σ', Σ'' and for every ω.

Now suppose, if possible, that η_0 is not in U, so that η_0 does not satisfy (15). Then the alternatives (26a), (26b) implies that

$$\omega \cdot \eta_0 > h_1(\omega) - \sum_{k>1} h_k(-\omega). \tag{27}$$

On the other hand it follows from (20) and from the absolute-uniform convergence of the series involved that

$$\omega \cdot \eta_0 < \sum_{k \leqq p} h_k(\omega) - \sum_{k>p} h_k(-\omega), \tag{28}$$

[9] Only at this point the treatment becomes essentially different from the one given by Kershner [4].

for a sufficiently large p and for every ω. Thus the alternatives (23a), (23b) and (7) imply the existence of an integer l such that

$$\sum_{k<l} h_k(\omega) - h_l(-\omega) < \omega\cdot\eta_0 + \sum_{k<l} h_k(-\omega) < \sum_{k<l} h_k(\omega) + h_l(\omega), \tag{29}$$

for every ω. On denoting by S^1, S^2 the convex hypersurfaces represented by

$$\begin{aligned} S^1:\quad \eta^1(\omega) &= \sum_{k<l} \eta_k(\omega); \qquad & h^1(\omega) &= \sum_{k<l} h_k(\omega) \\ S^2:\quad \eta^2(\omega) &= \eta_l(\omega); & h^2(\omega) &= h_l(\omega) \end{aligned}$$

it is clear from (17) and (7) that the condition corresponding to (17) in the case of S^1, S^2 is satisfied, so that the result obtained above in the case of two hypersurfaces may be applied to S^1, S^2. Thus, by (29) the convex hypersurface represented by

$$\eta = \eta_0 - \sum_{k>l} \eta_k(\omega)$$

is contained in the vectorial sum of S^1 and S^2. But this is in contradiction with the assumption that η_0 is not in T. This completes the proof of the following

THEOREM 1. *Let S_k $(k=1,2,\cdots)$, be a sequence of convex hypersurfaces the representations* (1) *of which satisfy* (10) *and the supporting functions* (2) *of which satisfy* (17). *Let S be the convex hypersurface represented by* (11), *let C be the closed convex set determined by S and let U be the (possibly empty) open convex subset of C determined by* (15). *Then the vectorial sum T of $S_1, S_2, \cdots$ is $C - U$.*

The supporting function $h_R(\omega)$ of the boundary R of U satisfies the inequality

$$h_R(\omega) \leqq h_1(\omega) - \sum_{k>1} h_k(-\omega),$$

in view of the definition (15) of U. Moreover, if $\eta = \eta_0$ is any point of R and $\omega = \omega_0$ is such that the equality sign holds in (18), then

$$h_R(\omega_0) = h_1(\omega_0) - \sum_{k>1} h_k(-\omega_0) \tag{30}$$

in view of the expression (2) of the supporting function of a hypersurface.

II. The hypersurfaces S_E. Let again (9) be a sequence of convex hypersurfaces and suppose, for simplicity, that the functions (1) defining (9) are univalued functions of ω, i. e. that any supporting hyperplane of S_k has exactly one point in common with S_k; and that no degenerate S_k occur, i. e. that in

(7), the equality sign is excluded. In what follows any sum of the form $\Sigma\,\eta_k(\omega)$ will denote the zero-vector if the set of subscripts k over which the summation is taken is empty. Let ϵ_k $(k = 1, 2, \cdots)$, be -1, 0 or $+1$ and let E be a symbol of the form

$$(31) \qquad E \equiv (\epsilon_1, \epsilon_2, \cdots).$$

For a given symbol (31), let S_E denote the locus represented by

$$(32) \qquad S_E\colon \quad \eta = \eta_E(\omega) = \sum_{\epsilon_k \neq 0} \eta_k(\epsilon_k\omega).$$

The following remarks concerning these loci are quite obvious:

(i) The symbol obtained from E by replacing every ϵ_k by $-\epsilon_k$ determines again the locus S_E. Thus all loci S_E are obtained if one normalizes E by the restriction that the first ϵ_k in E which is not 0 shall be 1.

(ii) The locus (32) which corresponds to the symbol (31) in which $\epsilon_k = 1$ for every k, is the convex hypersurface denoted in Section I by S.

(iii) Similarly the locus (32) which corresponds to any symbol E in which ϵ_k is either 0 or 1 for every k is a convex hypersurface with the supporting function

$$(33) \qquad h_E(\omega) = \sum_{\epsilon_k = 1} h_k(\omega)$$

(iv) If E is any symbol (31), let S_I, S_{II} be the convex hypersurfaces represented by

$$(34) \qquad S_I\colon \quad \eta = \eta_I(\omega) = \sum_{\epsilon_k = 1} \eta_k(\omega); \qquad S_{II}\colon \quad \eta = \eta_{II}(\omega) = \sum_{\epsilon_k = -1} \eta_k(\omega).$$

Then S_E may be represented in the form

$$(35) \qquad S_E\colon \quad \eta = \eta_E(\omega) = \eta_I(\omega) + \eta_{II}(-\omega).$$

(v) The set of all symbols (31) may be made into a topological space $\mathcal{E}$ as follows: If an integer n, and n numbers $\epsilon_1{}^0, \cdots, \epsilon_n{}^0$ equal to -1, 0, or $+1$ are given, the subset of $\mathcal{E}$ formed by those symbols E for which $\epsilon_i = \epsilon_i{}^0$, $(i = 1, \cdots, n)$, is said to be open in $\mathcal{E}$. Also any sum of open sets in $\mathcal{E}$ is said to be open in $\mathcal{E}$. It is well known that the resulting topological space is homeomorphic with a Cantor set (except, of course, if the number of given convex hypersurfaces is finite). From the assumption (10) and the definition (32), it follows that $\eta_E(\omega)$ is a continuous function of the pair of variables E, ω.

(vi) A symbol (31) is said to be of infinite length if $\epsilon_k \neq 0$ for every k. The subset of the topological space $\mathcal{E}$ consisting of all symbols of infinite

length clearly is compact and the S_E corresponding to these symbols are subsets of the vectorial sum T of all S_k. From the last remark in (v) it follows that the point set F formed by the points of all these S_E is a closed subset of T. The set F, which contains the boundary of T by Section I, will be called the *irregular set* of T. The set $G = T - F$ is an open set and will be called the *regular set* of T. It may be seen from examples that G may be empty and that G may be dense in T.

(vii) If a symbol of infinite length is such that $\epsilon_k = -1$ for at least one k, then S_E does not have a point in common with the exterior boundary of T. In fact for such an E, from (32) and (2), since the equality sign in (7) is excluded,

$$\omega \cdot \eta_E(\omega) = \sum_k \epsilon_k h_k(\epsilon_k \omega) < \sum_k h_k(\omega),$$

so that the statement follows from the remark made in connection with (8).

(viii) A symbol (31) is said to be of the finite length l if $\epsilon_k \neq 0$ or $\epsilon_k = 0$ according as $k \leqq l$ or $k > l$. Clearly if E is of length l, then S_E is contained in the vectorial sum T_l of $S_1, \cdots, S_l$. The irregular set F_l of T_l is formed by the 2^{l-1} curves S_E corresponding to symbols (31) of length l.

In what follows use will be made of the following assumption, which will be referred to as condition (*):

(*) *The hypersurface S_E which belongs to the symbol*

$$E = (+1, -1, -1, -1, \cdots) \tag{36}$$

is convex and forms the interior boundary R of the vectorial sum T of the S_k.

Concerning this condition the following remarks hold:

(ix) If condition (*) is satisfied, then the parameter representation $\eta_R(\omega)$ and the supporting function $h_R(\omega)$ of R are given, in view of (30), by

$$\eta_R(\omega) = \eta_1(\omega) + \sum_{k>1} \eta_k(-\omega); \qquad h_R(\omega) = h_1(\omega) - \sum_{k>1} h_k(\omega). \tag{37}$$

(x) If condition (*) is satisfied and E is any symbol (31) for which $\epsilon_1 = 1$, then S_E is a convex hypersurface with the supporting function

$$\sum_{\epsilon_k \neq 0} \epsilon_k h_k(\epsilon_k \omega). \tag{38}$$

In fact, by (32) and (37), the parameter representation of S_E may be brought into the form

$$\eta_E(\omega) = \eta_R(\omega) + \sum_{k>1,\, \epsilon_k=0} \{-\eta_k(-\omega)\} + \sum_{k>1,\, \epsilon_k=1} \{\eta_k(\omega) - \eta_k(-\omega)\}. \tag{39}$$

Since the separate terms in (39) are representations of convex hypersurfaces by (3), it follows, as in connection with (11) and (12), that (39) represents a convex hypersurface and that the supporting function of (39) is (38).

(xi) The following particular case of (x) is worth noting: If E is a symbol (31) of length l, then S_E is a convex hypersurface, moreover the hyperplane of normal direction ω through the point (32) of S_E is a supporting plane of S_E.

(xii) If the condition in (xi) holds for every l, then the S_E corresponding to any symbol E of infinite length is a convex hypersurface, even if condition (*) is not satisfied. This follows easily from the last statement under (v).

(xiii) If (*) is satisfied and a symbol (31) is such that $\epsilon_k \neq 0$ for every k and $\epsilon_k = 1$ for at least one $k > 1$, then the open convex region determined by the corresponding S_E contains the interior boundary of T. In fact if $h_E(\omega)$ is the supporting function of S_E, then by (39)

$$h_E(\omega) - h_R(\omega) = \sum_{k>1,\, \epsilon_k=1} \{h_k(\omega) + h_k(-\omega)\} > 0,$$

since the equality sign is excluded in (7), so that the statement follows from the remark in connection with (8).

III. A further consideration of the sets *F* and *G* [II (vi)]. In what follows use will be made of the following remark. If $\eta = \eta_1(\omega)$ is the representation (1) of a convex hypersurface S_1 and H represents a sphere of radius $r < \rho$ and centre at the origin together with its interior, then the vectorial sum of S_1 and H does not contain the interior of any hypersphere of radius ρ which does not meet S_1. This is clear since the distance of any part of this vectorial sum to the nearest point of S_1 is less than ρ.

Now suppose that (9) is a sequence of convex hypersurfaces satisfying (10), (17) and the conditions formulated at the beginning of Section II. Suppose in addition that the origin of the vector-space is in the interior of each of the curves S_k, i. e., that

$$(40) \qquad \omega \cdot \eta_k(\omega) = h_k(\omega) > 0$$

for every k and every ω, and also that the S_E corresponding to (9) have property II, (xi) for every l. Let V be a given component of the regular set G [II, (vi)] of the vectorial sum T of the S_k. Now there exists an integer m, which depends on V and has the following properties:

(i) if E is a symbol (31) such that the corresponding S_E contains a point of the boundary V, then the ϵ_k in E for which $k > m$ are all equal.

(ii) for every $n > m$, the regular set G_n of the vectorial sum T_n of $S_1, \cdots, S_n$ has a component V_n which contains V. Moreover V_n contains V_{n+1} for every $n > m$.

It will be shown first that if $\rho > 0$ is such that V contains the interior of a hypersphere of radius ρ, and m is such that

$$a_k < \tfrac{1}{2}\rho \text{ for every } k > m, \tag{41}$$

where a_k is defined in (10), then m satisfies (i). In fact let E be a symbol (31) such that $\epsilon_p = -1$, $\epsilon_q = 1$, where p and q are fixed numbers larger than l, then the loci S_+, S_E, S_- determined by

$$\eta = \eta_E(\omega) + \eta_p(\omega) - \eta_p(-\omega), \quad \eta = \eta_E(\omega), \quad \eta = \eta_E(\omega) - \eta_q(\omega) + \eta_q(-\omega)$$

are convex hypersurfaces contained in F [II, (vi)]. Now it is clear from (7) and (41) that the remark at the beginning of this Section may be applied. This shows that a region which contains the interior of a hypersphere of radius ρ and which does not meet S_E but has a boundary point in S_E, must necessarily meet either S_+ or S_-. Thus V cannot have a boundary point on S_E and so an integer m which satisfies (41) has property (i).

Next it will be shown that an integer m has not only property (i) but also property (ii) if the numbers a_k occurring in (10) satisfy not only (41) but also

$$\sum_{k>m} a_k < \tfrac{1}{4}\rho, \tag{42}$$

where ρ is again the radius of a sphere, the interior of which is in V. Let E be any symbol (31) of length $l > m$ and let E_+, E_- be the symbols obtained from E by replacing all ϵ_k for which $k > l$ (which are 0 by II, (viii)) by $+1, -1$ respectively. If $\eta_+(\omega)$, $\eta_-(\omega)$ are the parameter representations (32) of the hypersurfaces S_+, S_- which correspond to E_+, E_-, then it follows from (40), (42) and (10) that $\eta_+(\omega) - \eta_-(\omega)$ (which is by (32), (3) and II, (iii) the representation (1) of some convex hypersurface) satisfies

$$0 < \omega \cdot (\eta_+(\omega) - \eta_-(\omega)) < \tfrac{1}{2}\rho.$$

It follows from the remark at the beginning of this Section that if the interior of some sphere of radius ρ does not meet S_+ or S_-, then it does not meet S_E. Thus if S_E is of length $l > m$, then S_E does not meet V and accordingly the region V_l as defined in (ii) must exist if $l > m$. Obviously $V_l > V_{l+1}$, $l > m$, follows by the same argument, since no essential use has been made of the fact that the number of hypersurfaces in (9) is infinite.

Thus a number m exists which has both property (i) and property (ii). Of course the results of this Section hold if condition (*) is satisfied.

IV. The density of convolutions of distribution functions corresponding to convex curves. From now on convex curves S_k in the (x, y)-plane will be considered. Let the convex curve S_k be given by

$$S_k: \quad x = x_k(\theta), \qquad y = y_k(\theta) \tag{43}$$

where θ is an angular parameter and where exactly one point of S_k corresponds to each value of θ. It is well known that (43) determines a distribution function $\phi_k(E)$, where E is any Borel set in the (x, y)-plane, by the rule that $\phi_k(E)$ is the θ measure of the set of values of θ for which the point (43) of S_k is in E. Obviously S_k is the spectrum of ϕ_k. The convolution

$$\psi_n = \phi_1 * \phi_2 * \cdots * \phi_n, \qquad n > 1 \tag{44}$$

has as its spectrum the vectorial sum T_n of $S_1, \cdots, S_n$. On the assumption that (43) has a continuous derivative and that

$$x'_k(\theta)^2 + y'_k(\theta)^2 \neq 0 \text{ for every } \theta, \tag{45}$$

and finally that no S_k contains a line segment, it has been proved [10] that the convolution of at least two ϕ_k is absolutely continuous and that the convolution of at least four ϕ_k has a continuous density. If it is also assumed that the functions (43) are regular analytic, then it has been shown [10] that the density of the convolution of two ϕ_k is a regular analytic function of x and y on the regular set of the vectorial sum of the corresponding S_k (cf. (vi) in Section II) and that this density is not bounded in the vicinity of any point of the irregular set of the vectorial sum. This statement forms the special case where $n = 2$ of the following

THEOREM 2_n. *Let the convex curves S_k $(k = 1, 2, \cdots)$, in the (x, y)-plane be given by the regular analytic parameter representation* (43) *satisfying* (45) *and let the corresponding S_E have property* (xi) *of Section II for $l < n$. Then the density $\delta_n(x, y)$ of* (44) *is regular analytic on the regular set G_n of the vectorial sum T_n of $S_1, \cdots, S_n$.*

It will be assumed that the following statement has been proved:

($\S_n$) On the assumptions of Theorem 2_n, not only the statement of the Theorem holds but also: If P is a common point of exactly p distinct curves S_E, where E is of length n, and S^i $(i = 1, \cdots, p)$, denotes these curves, then a vicinity U of P may be found, and a regular analytic function $\lambda_i(x, y)$ in the complement of S^i in U $(i = 1, \cdots, p)$, such that for any point (x, y) in U,

[10] Van Kampen and Wintner [6], p. 103.

$$\delta_n(x, y) = \sum_{i=1}^{p} \lambda_i(x, y). \tag{46}$$

Needless to say $\lambda_i(x, y)$ consists in general of distinct regular functions in the parts into which U is divided by S^i. Thus (46) is to the effect that singularities of $\delta_n(x, y)$ at the several curves S_E are added at common points of distinct curves S_E. In case $n = 2$, Theorem 2_n and $(\S_n)$ are correct, the latter by (vii) of Section II. It will be shown that $(\S_n)$ implies $(\S_{n+1})$, thus completing the proof by complete induction of Theorem 2_n.

If the conditions of Theorem 2_{n+1} are satisfied, then clearly F_n consists of 2^{n-1} convex analytic curves S_E, so that any singularity of F_n is of the type described in $(\S_n)$.

Since $\psi_{n+1} = \psi_n * \phi_{n+1}$, one has for the density $\delta_{n+1}(x, y)$ of ψ_{n+1}

$$\delta_{n+1}(x, y) = \int \delta_n(x - \xi, y - \eta)\phi_{n+1}(d\xi, \eta E)$$

over the (ξ, η)-plane, or, by the definition of $\phi_{n+1}(E)$,

$$\delta_{n+1}(x, y) = \int \delta_n(x - x_{n+1}(\theta), \; y - y_{n+1}(\theta))\, d\theta \tag{47}$$

over all angles θ. For a fixed (x_0, y_0) let $P_1, \cdots, P_q$ be the distinct common points of F_n and the curve

$$x = x_0 - x_{n+1}(\theta), \qquad y = y_0 - y_{n+1}(\theta) \tag{48}$$

and let $I_1, \cdots, I_q$ be non-overlapping θ-intervals containing $P_1, \cdots, P_q$ such that the arcs of (48) corresponding to these intervals are contained in the vicinities corresponding to $P_1, \cdots, P_q$ by $(\S_n)$. Clearly the contribution to (47) of the θ-intervals, obtained by omitting $I_1, \cdots, I_q$ from the range of θ, is regular in a vicinity of (x_0, y_0). Suppose that (48) is not tangent to the curve S^i of F_n which passes through $P = P_j$. Then the contribution

$$\int_{I_j} \lambda_i(x - x_{n+1}(\theta), y - y_{n+1}(\theta))\, d \tag{49}$$

to (47) obviously is regular in a vicinity of (x_0, y_0) in view of $(\S_n)$. Now suppose that $(48)_j$ is tangent to the curve S^i of F_n at $P = P_j$. Then clearly (49) is regular in a vicinity of (x_0, y_0) except at those points x, y of this vicinity for which the curve corresponding to (48) still is tangent to S^i. Thus $(\S_{n+1})$ will follow if it is shown that to a point of tangency P_j of (48) and S^i there corresponds a point of one of the curves of F_{n+1} at (x_0, y_0). Now, if ω' is the value of ω corresponding to the point P_j of S^i, then the corre-

sponding point of S_{n+1} belongs to ω' or to $-\omega'$, so that the statement follows from the definition (32) of the curves S_E constituting F_{n+1}.

Remark. In Theorem 2_n it may be allowed that one or more of the curves S_E occurring degenerates into a single point. In that case "a curve is tangent to S_E" should be read "the curve passes through the point into which S_E degenerates."

Now let the convex curves S_k given in the z-plane be infinite in number and let (10) be satisfied. Then[11] the infinite convolution

$$\psi = \lim \psi_n = \phi_1 * \phi_2 * \phi_3 * \cdots \tag{50}$$

is convergent and has as its spectrum the vectorial sum T of the S_k, while

$$\delta(x, y) = \lim \delta_n(x, y), \tag{51}$$

where $\delta(x, y)$ is the density of (50).

THEOREM 3. *Let the curves S_k be given by the regular analytic representations* (43) *satisfying* (45) *and let the corresponding S_E have property* (xi) *of Section* II *for every l. Then the density* (51) *of* (50) *is regular analytic on the regular set G of the vectorial sum T of the S_k.*

In fact, let (x_0, y_0) be any point of G, let $3\rho > 0$ be less than the distance from (x_0, y_0) to the irregular set in T and let N be an integer such that

$$\sum_{n=N+1}^{\infty} a_n < \rho,$$

where the a_n are the numbers occurring in (10). Then clearly the circular disk with radius 2ρ and centre at (x_0, y_0) belongs to the regular set G_N of T_N and the spectrum of

$$\psi^N = \phi_{N+1} * \phi_{N+2} * \cdots \tag{52}$$

is within a distance ρ from the origin. If $\delta^N(x, y)$ is the density of ψ^N, then

$$\delta(x, y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \delta_N(x-\xi, y-\eta)\,\delta^N(\xi, \eta)\,d\xi\,d\eta \tag{53}$$

since $\phi = \psi_N * \psi^N$. This integral need only be taken over the spectrum of ψ^N, since $\delta^N(\xi, \eta) = 0$ unless (ξ, η) is in this spectrum. On the other hand if (x, y) is within a distance ρ of (x_0, y_0) and (ξ, η) is in the spectrum of ψ^N, then $(x-\xi, y-\eta)$ is within a circle of radius 2ρ and centre at (x_0, y_0).

[11] *Ibid.*, pp. 184-186.

Thus if (x, y) is in a circle of radius ρ and centre at (x_0, y_0), then the integrand of (53) is a regular analytic function of (x, y), so that $\delta(x, y)$ is a regular analytic function of (x, y). This completes the proof of Theorem 3.

Clearly in Theorem 3 the condition that II, (xi) is satisfied for every l may be replaced by the condition that the S_k satisfy the condition (*) introduced in II.

V. The logarithm of the Riemann zeta function. If p_m denotes the m-th prime number it is known[12] that for a fixed $\sigma > 1$, the asymptotic distribution function of the almost periodic function

$$(54)\qquad \begin{aligned} f_\sigma(t) &= -\log\zeta(\sigma + it) + \tfrac{1}{2}\log\zeta(2\sigma) \\ &= \sum_{m=1}^{\infty} [\log(1 - p_m^{-\sigma - it}) - \tfrac{1}{2}\log(1 - p_m^{-2\sigma})] \end{aligned}$$

may be represented as the infinite convolution

$$\phi^\sigma = \phi_1^\sigma * \phi_2^\sigma * \cdots,$$

where ϕ_m^σ denotes the distribution function which belongs to the convex curve (43) defined by

$$(55)\qquad S_m^\sigma:\quad x + iy = \log(1 - p_m^{-\sigma}e^{i\theta}) - \tfrac{1}{2}\log(1 - p_m^{-2\sigma})$$

or in other words by

$$(56)\qquad x = x_m^\sigma = r(\theta, p_m^{-\sigma}), \qquad y = y_m^\sigma = s(\theta, p_m^{-\sigma}),$$

where

$$(57)\qquad S_\rho:\quad x = r(\theta, \rho) = \tfrac{1}{2}\log\frac{1 - 2\rho\cos\theta + \rho^2}{1 - \rho^2}, \quad y = S(\theta, \rho) = \arctan\frac{\rho\sin\theta}{1 - \rho\cos\theta}$$

is the equation of a system of curves S_ρ depending continuously on a parameter ρ, it being understood that the arc tan takes only values between $-\frac{1}{2}\pi$ and $+\frac{1}{2}\pi$. For $0 < \rho < 1$, S_ρ represents a convex analytic curve satisfying (45) and having both axes $x = 0$, $y = 0$ as lines of symmetry.

Let $h(\rho, \omega)$ denote the supporting function of S_ρ, where now ω denotes not as in I, II, III, a variable unit vector in the (x, y)-plane, but the angle between such a vector and the positive x-axis. For reasons of symmetry it is sufficient to consider only the interval $0 \leqq \omega \leqq \frac{1}{2}\pi$. Clearly

$$(58)\qquad \begin{aligned} h(\rho, 0) &= \tfrac{1}{2}\log[(1 + p_m^{-\sigma})/(1 - p_m^{-\sigma})], \\ h(\rho, \tfrac{1}{2}\pi) &= \arcsin\rho^{-\sigma}(\leqq \tfrac{1}{2}\pi). \end{aligned}$$

[12] Van Kampen and Wintner [6], Section 7, where further references are given.

Now let $\rho_1, \rho_2, \cdots$ be a sequence of numbers such that $\rho_1 < 1$ and $0 < \rho_{n+1} < \rho_n$ and that $\sum_n h(\rho_n, \omega)$ converges uniformly in ω. It will be shown that the function

(59) $$f(\omega) = h(\rho_1, \omega) - \sum_{n>2} h(\rho_n, \omega)$$

has the following properties:

(60 a) $f(\omega) > 0$ for every ω, in case $h(\rho_1, \frac{1}{2}\pi) > \sum_{n>2} h(\rho_n, \frac{1}{2}\pi)$,

(60 b) $f(\omega) < 0$ for every ω, in case $h(\rho_1, 0) < \sum_{n>2} h(\rho_n, 0)$,

(60 c) $f(\omega_0) = 0$, for an $\omega = \omega_0$, $f(\omega) > 0$ if $0 \leqq \omega < \omega_0$, $f(\omega) < 0$ if $\omega_0 < \omega \leqq \frac{1}{2}\pi$, in case $h(\rho_1, \frac{1}{2}\pi) \leqq \sum_{n>2} h(\rho_n, \frac{1}{2}\pi)$ and $h(\rho_1, 0) \geqq \sum_{n>2} h(\rho_n, 0)$.

Clearly (60 a), (60 b), (60 c) follow from the following statement

(61) $$\frac{h(\rho_1, \omega_1)}{h(\rho_1, \omega_2)} < \frac{h(\rho_2, \omega_1)}{h(\rho_2, \omega_2)} \text{ if } 0 < \rho_1 < \rho_2 < 1,\ 0 \leqq \omega_1 < \omega_2 \leqq \tfrac{1}{2}\pi.$$

For the proof of (61) use will be made of the following properties of the curves S^ρ: [13]

Let S_ρ denote the curve similar to S^ρ with respect to the origin in ratio $1/(\arcsin r)$, so that, by (58), the curves S_ρ have the points $(0, \pm 1)$ in common for $0 < \rho < 1$. Then S_{ρ_1} is, except for those two points, in the interior of S_{ρ_2}, whenever $0 < \rho_1 < \rho_2 < 1$. Moreover, if $\alpha(\rho, s)$ denotes, for $0 < s < 1$, the angle between the line $y = s$ and the normal of S_ρ at the intersection of this line and S_ρ, then $\alpha(\rho, s)$ is an increasing function of ρ in $0 < \rho < 1$.

In proving (61) it is clearly admissible to replace $h(\rho, \omega)$ by the supporting function of any convex curve S'_ρ similar to S^ρ with respect to the origin. On choosing, for fixed values ρ_1, ρ_2 $(\rho_1 < \rho_2)$ and ω_1, the curves S'_ρ in such a way that $h(\rho_1, \omega_1) = h(\rho_2, \omega_2)$, it is clear from the geometrical properties of S_ρ mentioned above that $h(\rho_2, \omega_2) < h(\rho_1, \omega_1)$ if $\omega_2 > \omega_1$ is sufficiently near to ω_1. Thus $h(\rho_1, \omega)/h(\rho_2, \omega)$ is an increasing function of ω if $0 < \rho_1 < \rho_2 < 1$ and (61) is proved.

Let σ^k, σ_k denote, for any positive integer k, the obviously unique numbers such that

(62 a) $$\arcsin p_k^{-\sigma} = \sum_{m>k} \arcsin p_m^{-\sigma} \text{ if } \sigma = \sigma^k > 1$$

(62 b) $$\tfrac{1}{2}\log\left[(1 + p_k^{-\sigma})/(1 - p_k^{-\sigma})\right] = \sum_{m>k} \tfrac{1}{2}\log\left[(1 + p_m^{-\sigma})/(1 - p_m^{-\sigma})\right] \text{ if } \sigma = \sigma_k > 1$$

[13] Bohr and Jessen [1]. In this paper (61) is proved for the case $\omega_2 = \frac{1}{2}\pi$. The general proof given below is an extension of the proof given there.

and let $\bar{\sigma}$ denote the number such that

$$(62\text{ c}) \qquad p_1^{-\sigma} = \sum_{m>1} p_m^{-\sigma}, \text{ if } \sigma = \bar{\sigma} > 1.$$

Then it is clear from (60 a), (60 b), (60 c) that $\sigma^k > \sigma_k$ for every $k > 0$, while the numerical values $\sigma^1 = 1.764\ldots$ and $\bar{\sigma} = 1.778\ldots$ show that $\sigma_1 < \sigma^1 < \bar{\sigma}$. Clearly for sufficiently large k, $\sigma^k > \sigma_k > \bar{\sigma}$.

Consider now the case [14] where for some integer l,

$$(63) \qquad \sigma > \operatorname{Max}(\bar{\sigma}^{(1)}, \cdots, \sigma^{(l)}).$$

Let E_I and E_{II} be the symbols (31) obtained from a given symbol of length $l-1$ by replacing ϵ_l by $+1$, ϵ_k by -1, $k > l$, and ϵ_l by -1, ϵ_k by $+1$, $k > l$, respectively. Let S_I, S_{II} be the corresponding curves (32), $\eta_I(\omega)$, $\eta_{II}(\omega)$ their parametric representations and $h_I(\omega)$, $h_{II}(\omega)$ their supporting functions. It is known [15] that, whenever $\sigma \geqq \bar{\sigma}$, the vectorial sum of the S_m^σ satisfies condition (*), so that, by II (x), the curves S_E corresponding to the S_m^σ are convex curves and Theorem 3 is applicable. Now the functions h_I, h_{II} satisfy, by (32) and the definition of $h(\rho, \omega)$, the equality

$$(64) \qquad h_I(\omega) - h_{II}(\omega) = 2h(p_l^{-\sigma}, \omega) - 2\sum_{m>l} h(p_m^{-\sigma}, \omega).$$

Thus it follows from (60 a), (58) and the definition (62 a) of σ^k that $h_I(\omega) - h_{II}(\omega) > 0$. Hence S_I and S_{II} are exterior and interior boundary of a ringshaped subset of the vectorial sum of all S_m^σ. In a similar way it may be shown that in view of (63) no S_E corresponding to a symbol E of infinite length does enter the ring shaped region determined by S_I and S_{II}. Thus this ring shaped region is by Theorem 3 a region of regular analyticity for the density of ϕ^σ.

A region of regular analyticity as found above does not disappear abruptly if σ does not satisfy (63) but is very near to values satisfying (63). In fact, since all curves S_E depend on σ in a uniformly continuous way, if σ is restricted to a bounded interval with positive endpoints, it is clear that the ring shaped region described above remains ring shaped for certain σ if $\sigma > \operatorname{Max}(\bar{\sigma}, \sigma^l)$ remains satisfied, but σ becomes less than σ^k for some $k = 1, \cdots, l-1$.[16] It will be shown below that the ring shaped region constructed above is replaced by a pair of crescent shaped regions if σ satisfies

---

[14] This case has already been considered by van Kampen and Wintner [6], Section 7.

[15] Bohr and Jessen [1]; Kershner [5].

[16] It seems to require elaborate numerical calculations to decide whether or not the sequences $\{\sigma^k\}$, $\{\sigma_k\}$ are monotone.

$$(65) \qquad \sigma^k > \sigma > \operatorname{Max}(\bar{\sigma}, \sigma^1, \cdots, \sigma^{l-1}, \sigma_l)$$

In fact, if (65) is satisfied, then the difference (64) of $h_I(\omega)$ and $h_{II}(\omega)$ is a function satisfying (60 c), so that the curves S_I and S_{II} determine four crescent shaped regions, two of which are halved by the x-axis and two of which are halved by the y-axis. It is clear from (60 c) that the first two regions have S_I as exterior and S_{II} as interior boundary, while the opposite is true for the last two regions. Since S_I is in the interior of the curves S_E corresponding to symbols E of infinite length adjacent to E_I, and S_{II} is exterior to the curves S_E corresponding to symbols E of infinite length adjacent[17] to E_{II}, it is clear that the last two regions are not contained in the regular set of the vectorial sum of all $S_m{}^\sigma$. On the other hand, it may be inferred easily from (65) that no S_E corresponding to symbols E of infinite length has a point in common with the two crescent shaped regions which are halved by the x-axis. Thus the latter two regions are regions of regular analyticity of the density of ϕ^σ by Theorem 3. It is clear that the two regions just described decrease in size as σ decreases from σ^l to σ_l and disappear completely when σ becomes equal to σ_l. The two regions may, of course, cease to be regions of regular analyticity of the density of ϕ^σ for some $\sigma > \sigma_l$ if for instance $\sigma^{l-1} > \sigma_l$, a case which probably occurs for sufficiently large l.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] H. Bohr and B. Jessen, "On the distribution of the values of the Riemann zeta function," *American Journal of Mathematics*, vol. 58 (1936), pp. 35-44.

[2] E. K. Haviland, "On the addition of convex curves in Bohr's theory of Dirichlet series," *ibid.*, vol. 55 (1933), pp. 332-334.

[3] B. Jessen and A. Wintner, "Distribution functions and the Riemann zeta function," *Transactions of the American Mathematical Society*, vol. 38 (1935), pp. 48-88.

[4] R. B. Kershner, "On the addition of convex curves," *American Journal of Mathematics*, vol. 58 (1936), pp. 737-746.

[5] R. B. Kershner, "On the values of the Riemann zeta function on fixed lines $\sigma > 1$," *ibid.*, vol. 59 (1937), pp. 171-178.

[6] E. R. van Kampen and A. Wintner, "Convolutions of distribution functions on convex curves and the Riemann zeta function," *ibid.*, vol. 59 (1937), pp. 175-204.

[17] For the meaning of words like "adjacent" as applied to symbols (31) compare II, (v).

ON THE PARTIAL SUMS OF CERTAIN FOURIER SERIES.*

By OTTO SZÁSZ.

1. Suppose $f(x)$ is real-valued, periodic of period 2π, and Lebesgue integrable over $(-\pi, \pi)$. Denote its Fourier series by

$$f(x) \sim a_0/2 + \sum_{\nu=1}^{\infty} (a_\nu \cos \nu x + b_\nu \sin \nu x), \tag{1}$$

and let its partial sums be

$$s_0 = a_0/2, \; s_n = s_n(x) = a_0/2 + \sum_{\nu=1}^{n} (a_\nu \cos \nu x + b_\nu \sin \nu x), \; (n = 1, 2, \cdots).$$

In 1932, Paley [3] proved the two theorems which follow:

I. *Suppose*

$$|f(x)| \leq M \text{ in } (-\pi, \pi), \tag{2}$$

$$a_0 \geq 0, \qquad a_n \geq 0, \qquad b_n \geq 0, \qquad (n = 1, 2, \cdots), \tag{3}$$

then

$$|s_n(x)| \leq 10M, \qquad -\pi \leq x \leq \pi, \qquad (n = 0, 1, 2, \cdots). \tag{4}$$

II. *If in addition $f(x)$ is continuous, then its Fourier series* (1) *converges to $f(x)$ uniformly for all x.*

For Theorem I Fejér [1] gave a simpler proof and replaced the constant 10 by 4, so that

$$|s_n(x)| \leq 4M. \tag{5}$$

I give in (§ 2) another elementary proof for both theorems simultaneously. At the same time I replace (5) by (§§ 3-4)

$$|s_n(x)| \leq M(2 + \alpha/\sin^2 \alpha) < 3.38M, \tag{6}$$

where α is the unique root of the equation $2\alpha = \tan \alpha$ in the interval $0 < \alpha < \pi/2$.

If β is the least number such that (2) and (3) imply

$$|s_n(x)| \leqq \beta M, \qquad -\pi \leq x \leq \pi, \qquad (n = 0, 1, 2, \cdots),$$

* Received April 26, 1937.

then, by (6),

$$\beta \leqq 2 + \alpha/\sin^2 \alpha < 3.38.$$

On the other hand if $f(x) = \pi - x/2$, $0 < x < \pi$, $f(-x) = -f(x)$, then

$$f(x) = \sum_1^\infty \frac{\sin \nu x}{\nu},$$

$$M = \pi/2, \text{ and } s_n(\pi/n+1) \uparrow \int_0^\pi \frac{\sin x}{x}\, dx = 1.851 \cdots;$$

hence

$$\beta \geqq \frac{2}{\pi}\int_0^\pi \frac{\sin x}{x}\, dx > 1.17.$$

Thus β lies somewhere between 1.17 and 3. 38.

In §§ 5-6, I consider generalized trigonometric series, improving on a theorem by M. Fekete [2].

2. On putting

$$\frac{f(x)+f(-x)}{2} = \phi(x), \quad \int_0^x \phi(t)\, dt = \phi_1(x), \quad \int_0^x \phi_1(t)\, dt = \phi_2(x),$$

we have

$$\phi(x) \sim \frac{a_0}{2} + \sum_{\nu=1}^\infty a_\nu \cos \nu x$$

$$\phi_1(x) = \tfrac{1}{2}a_0 x + \sum_{\nu=1}^\infty \frac{a_\nu}{\nu} \sin \nu x$$

$$\phi_2(x) = \tfrac{1}{4}a_0 x^2 + \sum_{\nu=1}^\infty \frac{a_\nu}{\nu^2}(1-\cos \nu x) = \tfrac{1}{4}\, a_0 x^2 + 2\sum_{\nu=1}^\infty \frac{a_\nu}{\nu^2} \sin^2 \tfrac{1}{2}\nu x$$

$$\geq a_0 \left(\frac{x}{2}\right)^2 + 2\left(\frac{x}{2}\right)^2 \sum_{\nu=1}^n a_\nu \left(\frac{\sin \frac{1}{2}\nu x}{\frac{1}{2}\nu x}\right)^2, \qquad (n = 1, 2, \cdots),$$

if $a_n \geq 0$, $(n = 1, 2, \cdots)$. We now assume in addition, instead of (2),

$$\phi(x) \leqq M \text{ in } (-\pi, \pi).$$

Hence, a fortiori, $\phi_2(x) \leq x^2 M/2$ in $0 < x < \pi$, and

$$\frac{a_0}{2} + \sum_{\nu=1}^n a_\nu \left(\frac{\sin \frac{1}{2}\nu x}{\frac{1}{2}\nu x}\right)^2 \leq M, \qquad (n = 1, 2, 3, \cdots). \tag{7}$$

Letting now $x \downarrow 0$, (7) takes the form

$$\frac{a_0}{2} + \sum_{\nu=1}^n a_\nu \leq M, \qquad (n = 1, 2, \cdots). \tag{8}$$

Hence Σa_ν converges, and $\Sigma a_\nu \cos \nu x$ converges uniformly in any interval. Moreover, by (8),

$$\sum_{\nu=1}^{n} a_\nu \leq M - \frac{a_0}{2}; \text{ whence } \sum_{\nu=1}^{n} a_\nu \mid \cos \nu x \mid \leq M - \frac{a_0}{2},$$

and

$$(9) \qquad \frac{a_0}{2} + \sum_{\nu=1}^{n} a_\nu \cos \nu x \leq \frac{a_0}{2} + \sum_{1}^{n} a_\nu \mid \cos \nu x \mid \leq M.$$

Furthermore, let $\phi(x) \geqq -M$; then from $\frac{a_0}{2} = \frac{1}{\pi}\int_0^\pi \phi(x)\,dx$ it follows that $-2M \leqq a_0 \leq 2M$. Also by (8)

$$(10) \qquad \tfrac{1}{2}a_0 + \sum_{1}^{n} a_\nu \cos \nu x \geq -\sum_{1}^{n} a_\nu \geq \tfrac{1}{2}a_0 - M \geq -M.$$

On putting

$$\frac{f(x) - f(-x)}{2} = \psi(x), \qquad \int_0^x \psi(t)\,dt = \psi_1(x),$$

we have

$$\psi(x) \sim \sum_{1}^{\infty} b_\nu \sin \nu x,$$

$$\frac{1}{x}\psi_1(x) = 2\sum_{1}^{\infty} b_\nu \frac{\sin^2 \frac{1}{2}\nu x}{\nu x} \geq 2\sum_{1}^{n} b_\nu \frac{\sin^2 \frac{1}{2}\nu x}{\nu x} = \frac{x}{2}\sum_{1}^{n} \nu b_\nu \left(\frac{\sin \frac{1}{2}\nu x}{\frac{1}{2}\nu x}\right)^2.$$

Since $\sin u/u$ is monotonic decreasing in $0 < u < \pi/2$, this becomes for $x = \pi/n$

$$\frac{4}{\pi^2} \cdot \frac{\pi}{2n}\sum_{1}^{n} \nu b_\nu \leq \frac{n}{\pi}\psi_1\left(\frac{\pi}{n}\right), \qquad (n = 1, 2, 3, \cdots),$$

or

$$(11) \qquad \frac{2}{n\pi}\sum_{1}^{n} \nu b_\nu \leq \frac{n}{\pi}\psi_1\left(\frac{\pi}{n}\right), \qquad (n = 1, 2, 3, \cdots),$$

If now $\psi(x) \leq M$ in $\mid x \mid < \pi$, and a fortiori $\psi_1(x)/x \leq M$, $0 < x < \pi$, then (11) leads to

$$(12) \qquad \frac{2}{n\pi}\sum_{1}^{n} \nu b_\nu \leq M, \qquad \frac{1}{n}\sum_{1}^{n} \nu b_\nu \leq \frac{\pi}{2}M, \qquad (n = 1, 2, 3, \cdots).$$

If, in addition, $\psi(x)$ is continuous at $x = 0$, then $\psi_1(x)/x \to 0$ for $x \downarrow 0$, and from (11) we derive

$$(13) \qquad \sum_{1}^{n} \nu b_\nu = o(n), \qquad n \to \infty.$$

Using the identity

$$(14) \qquad U_n = \frac{1}{n}\sum_{1}^{n-1} U_\nu + \frac{1}{n}\sum_{1}^{n} \nu u_\nu, \qquad U_n = \sum_{1}^{n} u_\nu, \qquad (n = 2, 3, \cdots),$$

where we put $u_n = b_n \sin nx$, Fejér's theorem on the arithmetic means and (12) lead to

$$(15) \qquad \left| \sum_1^n b_\nu \sin \nu x \right| \leq M + \frac{\pi}{2} M = \left(1 + \frac{\pi}{2}\right) M;$$

(9), (10) and (15) yield Theorem I with the smaller constant $2 + \pi/2$ instead of 4 in the inequality (5).

Moreover, if $f(x)$ is continuous throughout, Theorem II follows from (13) and (14).

3. We shall generalize the assumptions on $\phi(x)$ and $\psi(x)$, and improve upon the constant in (12). We first prove

THEOREM 1. *Let* $a_\nu \geq 0$, $(\nu = 0, 1, 2, \cdots)$, *and* $\sum_1^\infty \nu^{-2} a_\nu < \infty$; *then*

$$(16) \qquad \frac{a_0}{2} + \sum_{\nu=1}^\infty \left(\frac{\sin \nu x}{\nu x}\right)^2 a_\nu = R_1(x)$$

exists for $x > 0$. *Suppose moreover*

$$\liminf_{x \to 0} R_1(x) = M,$$

then

$$\tfrac{1}{2} a_0 \leq M, \quad \left| \tfrac{1}{2} a_0 + \sum_1^n a_\nu \cos \nu x \right| \leq M, \quad |x| < \pi, \qquad (n = 1, 2, \cdots),$$

and $\sum a_\nu \cos \nu x$ *converges uniformly on the real axis.*

For the proof, note that (16) implies

$$\frac{a_0}{2} + \sum_1^n \left(\frac{\sin \nu x}{\nu x}\right)^2 a_\nu \leq R_1(x), \qquad (n = 1, 2, 3, \cdots);$$

let $x_k \downarrow 0$ and $\lim_{k \to \infty} R_1(x_k) = M$; then

$$\lim_{k \to \infty} \left\{ \frac{a_0}{2} + \sum_1^n \left(\frac{\sin \nu x_k}{\nu x_k}\right)^2 a_\nu \right\} \leq \lim_{k \to \infty} R_1(x_k) = M.$$

Hence,

$$\frac{a_0}{2} + \sum_1^n a_\nu \leq M, \quad (n = 1, 2, \cdots), \text{ and } \tfrac{1}{2} a_0 \leqq M.$$

The theorem follows at once from this.

If in particular $\phi(x) \sim \frac{1}{2} a_0 + \sum_1^\infty a_\nu \cos \nu x$, then

$$R_1(\tfrac{1}{2} x) = \frac{2}{x^2} \int_0^x \phi(t)(x - t)\, dt,$$

and we have the

COROLLARY. *If*

$$\phi(x) \sim \tfrac{1}{2}a_0 + \sum_1^\infty a_\nu \cos \nu x, \qquad a_\nu \geq 0,$$

and

$$\liminf_{x\to\infty} \frac{2}{x^2}\int_0^x \phi(t)\,(x-t)\,dt = M$$

then $\sum_1^\infty a_\nu < \infty$ *and* $\tfrac{1}{2}a_0 + \sum_1^\infty a_\nu \leq M.$

We call $u_0 + \sum_1^\infty \left(\frac{\sin \nu x}{\nu x}\right)^2 u_\nu$ the Riemannian mean (

sponding to the series $\sum_1^\infty u_\nu$.

4. We next prove

THEOREM 2. *Let* $b_\nu \geqq 0$, $(\nu = 1, 2, 3, \cdots)$, *and*

$$(17) \qquad \frac{2x}{\pi}\sum_1^\infty \nu b_\nu \left(\frac{\sin \nu x}{\nu x}\right)^2 = R_2(x)$$

exists for $x > 0$. *Suppose, in addition,* $R_2(x) \leqq 2$
$0 < x \leq \delta$; *then*

$$(18) \qquad \sum_1^n \nu b_\nu \leq \frac{\alpha}{\sin^2 \alpha} Mn < 1.38 Mn \quad \textit{for} \quad n$$

where α *is the unique root of the equation* $2\alpha = \tan \alpha$ i

We have from (17)

$$R_2(x) \geq \frac{2x}{\pi}\sum_1^n \nu b_\nu \left(\frac{\sin \nu x}{\nu x}\right)^2 \geq \frac{2x}{\pi}\left(\frac{\sin nx}{nx}\right.$$

for $0 < nx \leq \pi/2$; or

$$(19) \quad \frac{1}{n}\sum_1^n \nu b_\nu \leq \frac{\pi}{2}\;\frac{nx}{\sin^2 nx} R_2(x) \leq M\,\frac{nx}{\sin^2 nx} \leq M\,\frac{\alpha}{\sin}$$

where $\alpha/\sin^2\alpha$ is the minimum of $t/\sin^2 t$ in $0 < t < \pi$
to be the unique root of the equation $2\alpha = \tan\alpha$ in $0 <$
of Functions" by Jahnke and Emde (2nd edition,
$1.16 < \alpha < 1.17 = \alpha_1$, and $\sin \alpha_1/\alpha_1 = 0.7870\cdots$.
calculation

$$\frac{\alpha_0}{\sin^2 \alpha_0} < \frac{\alpha_1}{\sin^2 \alpha_1} < 1.38.$$

This proves the theorem.

If in particular $\psi(x) \sim \sum_1^\infty b_\nu \sin \nu x$ and $|\psi(x)| \leqq M$, then relations (18) and (14), with $u_\nu = b_\nu \sin \nu x$, give

$$\left|\sum_1^n b_\nu \sin \nu x\right| \leq M\left(1 + \frac{\alpha}{\sin^2 \alpha}\right) < 2.38M.$$

Furthermore

$$\frac{1}{x}\int_0^x \psi(t)\,dt = \sum_1^n b_\nu \frac{1 - \cos \nu x}{\nu x} = \frac{x}{2}\sum_1^\infty \nu b_\nu \left(\frac{\sin \frac{1}{2}\nu x}{\frac{1}{2}\nu x}\right)^2 = \frac{\pi}{2} R_2\left(\frac{x}{2}\right);$$

and, from (19),

$$\frac{1}{n}\sum_1^n \nu b_\nu \leq \frac{\alpha}{\sin^2 \alpha} \cdot \frac{1}{x}\int_0^x \psi(t)\,dt, \qquad 0 < nx \leq \pi;$$

in particular, if $\frac{1}{x}\int_0^x \psi(t)\,dt \to 0$ as $x \to 0$, then

$$\frac{1}{n}\sum_1^n \nu b_\nu \to 0 \text{ for } n \to \infty.$$

We call $\frac{2x}{\pi}\left\{ s_0 + \sum_1^\infty \left(\frac{\sin \nu x}{\nu x}\right)^2 s_\nu \right\}$ the Riemannian mean of the second kind corresponding to the sequence $\{s_n\}$, or to the series $\sum_0^\infty u_\nu$ with $\sum_0^n u_\nu = s_n$.

5. We now pass to generalized trigonometric series and to almost periodic functions. The most general result in the case of positive coefficients is due to M. Fekete [2].

Let $\phi(x)$ be a measurable real function of the real variable x,

$$\sup_x \frac{1}{x}\int_0^x \phi^2(t)\,dt < \infty,$$

and let $\phi(-x) = \phi(x)$. Suppose that $\phi(x) \leqq U$, and that

$$\lim_{x\to\infty} \frac{1}{x}\int_0^x \phi(t) \cos \lambda t\,dt = a(\lambda) = M\{\phi(t) \cos \lambda t\} \tag{20}$$

exists for all $\lambda \geqq 0$. It is easy to show that $a(\lambda)$ vanishes, except at an enumerable set of λ-values. Denote, in a certain order, by $\lambda_1, \lambda_2, \cdots$ those λ for which $a(\lambda) \neq 0$; we call the series

$$\sum_1^\infty a_n \cos \lambda_n x \sim \phi(x), \tag{21}$$

where $a_0 = a(0)$, $a_n = 2a(\lambda_n)$, in the Fourier expansion of $\phi(x)$.

Let $\rho_1, \rho_2, \cdots$ denote a subsequence of $\{\lambda_n\}$, consisting of linearly independent numbers; that is, no equation of the form $\sum_1^n r_\nu \rho_\nu = 0$ holds, where the r_ν are rational and not all zero. Denote by $\{\mu_n\}$ the subsequence of those λ_n, which can be represented in the form

$$\lambda_n = \sum_{\nu=1}^{h} r_\nu \rho_\nu, \qquad h = h(n); \qquad r_\nu = r_\nu(n) \text{ rational.}$$

Let $0 < \omega < \Omega$ denote given positive numbers. With this notation (with the restriction $|\phi(x)| \leq U$) Fekete proved the theorem:

Suppose that

$$a(\mu_n) \geq 0 \quad \textit{for} \quad 0 \leq \mu_n < \Omega;$$

then the "partial sum"

$$a_0 + 2 \sum_{\mu_n < \omega} a(\mu_n) \cos \mu_n x, \qquad \omega < \Omega \tag{22}$$

of the Fourier expansion (21) *converges absolutely in* $-\infty < x < \infty$; *its sum function* $\phi_\omega(x)$, *an almost periodic function whose Fourier expansion is identical with* (22), *satisfies the inequality*

$$|\phi_\omega(x)| \leq \frac{U}{1 - \omega/\Omega}. \tag{23}$$

For the proof, we associate with (21) the "Fejér-polynomials" $\sum_q(x)$, defined, for $q \geq 1$, by

$$\begin{aligned}\Sigma_q(x) = M \Big\{ \phi(t) \sum_{\nu_1=-P}^{P} \sum_{\nu_2=-P}^{P} \cdots \sum_{\nu_q=-P}^{P} \left(1 - \frac{|\nu_1|}{P}\right) \\ \cdots \left(1 - \frac{|\nu_q|}{P}\right) \cos t \, \frac{\nu_1\rho_1 + \cdots + \nu_q\rho_q}{Q} \cos x \, \frac{\nu_1\rho_1 + \cdots + \nu_q\rho_q}{Q} \Big\},\end{aligned} \tag{24}$$

where $Q = q!$, $P = qQ = q \cdot q!$. If the sequence (ρ_ν) consists of q_0 terms only, we put $\rho_\nu = 0$ for $\nu > q_0$. By virtue of (20) and the linear independence of the ρ_ν's, we can write (24) in the form

$$\Sigma_q(x) = a_0 + \sum k_n^{(q)} a_n \cos \lambda_n x, \tag{25}$$

where

$$k_n^{(q)} = \prod_{a=1}^{q} \left(1 - \frac{|\nu_a|}{P}\right), \text{ whenever } 0 < \lambda_n = \frac{1}{Q} \sum_{a=1}^{q} \nu_a \rho_a, \tag{26}$$

the ν_a being integers with $|\nu_a| < P$, while $k_n^{(q)} = 0$ for other values. Obviously $0 \leq k_n^{(q)} < 1$. We shall prove

$$\lim_{q \to \infty} k_n^{(q)} = 1, \text{ whenever } \lambda_n \subset \{\mu_\nu\}. \tag{27}$$

In fact, we then have

$$\lambda_n = \sum_{\alpha=1}^{h} r_\alpha \rho_\alpha = \sum_{\alpha=1}^{h} \frac{u_\alpha}{v_\alpha} \rho_\alpha, \qquad u_\alpha \text{ and } v_\alpha > 0, \text{ integers.}$$

Hence, for

$$q > \max(v_1, \cdots, v_h, h, |r_1|, \cdots, |r_h|),$$

$$\lambda_n = \frac{\nu_1}{Q}\rho_1 + \cdots + \frac{\nu_h}{Q}\rho_h + \frac{0}{Q}\rho_{h+1} + \cdots + \frac{0}{Q}\rho_q,$$

where $\nu_\alpha = Qr_\alpha$ for $1 \leq \alpha \leq h$. Consequently, from (26),

$$k_n^{(q)} = \prod_{\alpha=1}^{h}\left(1 - \frac{Q\,|\,r_\alpha\,|}{p}\right) = \prod_{\alpha=1}^{h}\left(1 - \frac{|\,r_\alpha\,|}{q}\right) \uparrow 1, \text{ for } q \to \infty.$$

Finally we prove

(28) $$\Sigma_q(x) \leq U.$$

This follows from

$$\begin{aligned}\Sigma_q(x) = M\Big\{\phi(t) \sum_{\nu_1,\ldots,\nu_q}^{-P,P} \prod_{1}^{q}\left(1 - \frac{|\,\nu_\alpha\,|}{P}\right) \tfrac{1}{4}\,[\exp(i(t+x)\sum_1^q \nu_\alpha\rho_\alpha/Q) \\ + \exp(-i(t+x)\sum_1^q \nu_\alpha\rho_\alpha/Q) + \exp(i(t-x)\sum_1^q \nu_\alpha\rho_\alpha/Q) \\ + \exp(-i(t-x)\sum_1^q \nu_\alpha\rho_\alpha/Q)]\Big\} \\ = M\{\phi(t)\tfrac{1}{2}[K_q(t+x) + K_q(t-x)]\},\end{aligned}$$

where

$$\begin{aligned}K_q(t) &= \sum_{\nu_1,\ldots,\nu_q}^{-P,P} \prod_{1}^{q}\left(1 - \frac{|\,\nu_\alpha\,|}{P}\right)\exp(it\sum_1^q \nu_\alpha\rho_\alpha/Q) \\ &= \prod_{\alpha=1}^{q} \sum_{\nu_\alpha=-P}^{P}\left(1 - \frac{|\,\nu_\alpha\,|}{P}\right)\exp(i\nu_\alpha\rho_\alpha t/Q) \\ &= \prod_{\alpha=1}^{q} \frac{1}{P}\left[\frac{\sin\dfrac{P}{2}\dfrac{\rho_\alpha}{Q}t}{\sin\dfrac{\rho_\alpha}{2Q}}\right]^2.\end{aligned}$$

From this we see that

$$\Sigma_q(x) \leq UM\{\tfrac{1}{2}[K_q(t+x) + K_q(t-x)]\} = U,$$

and, in particular,

$$\Sigma_q(0) = a_0 + \sum k_n^{(q)} a_n \leq U.$$

We next define for any positive μ the "Riesz-mean"

$$R_\mu(x, q) = a_0 + \sum_{0<\lambda_n<\mu}\left(1 - \frac{\lambda_n}{\mu}\right) k_n^{(q)} a_n \cos \lambda_n x$$

associated with the polynomial (25). Applying the formulas

$$\frac{2}{\pi}\int_0^\infty \cos 2kt \,\frac{\sin^2 \mu t}{\mu t^2}\, dt = \begin{cases} 1 - k/\mu & \text{if} \quad 0 \leq k < \mu \\ 0 & \text{if} \quad 0 < \mu \leq k \end{cases} \tag{29}$$

and using (25), we obtain

$$R_\mu(x, q) = \frac{2}{\pi}\int_0^\infty \tfrac{1}{2}[\Sigma_q(x + 2t) + \Sigma_q(x - 2t)]\, \frac{\sin^2 \mu t}{\mu t^2}\, dt.$$

From this and (28) we find

$$R_\mu(x, q) \leq \frac{2}{\pi}\, U \int_0^\infty \frac{\sin^2 \mu t}{\mu t^2}\, dt = U;$$

and, in particular,

$$R_\mu(0, q) = a_0 + \sum_{0 < \lambda_n < \mu} \left(1 - \frac{\lambda_n}{\mu}\right) k_n^{(q)} a_n \leq U.$$

Since every term is positive if $\mu < \Omega$, we get for an arbitrary N

$$a_0 + \sum_{0 < \lambda_n < \mu}^{n \leq N} \left(1 - \frac{\lambda_n}{\mu}\right) k_n^{(q)} a_n \leq U.$$

Now passing to the limit $q \to \infty$,

$$a_0 + 2 \sum_{0 < \mu_m < \mu}^{m \leq M} \left(1 - \frac{\mu_m}{\mu}\right) a(\mu_m) \leq U, \qquad M = M(N),$$

and, a fortiori for $0 < \omega < \mu < \Omega$,

$$a_0 + 2 \sum_{0 < \mu_m < \omega}^{m \leq M} \left(1 - \frac{\mu_m}{\mu}\right) a(\mu_m) \leq U.$$

Here $1 - \mu_m/\mu > 1 - \omega/\mu$; hence,

$$a_0 + 2\left(1 - \frac{\omega}{\mu}\right) \sum_{0 < \mu_m < \omega}^{m \leq M} a(\mu_m) \leq U,$$

and for $\mu \to \Omega$

$$a_0 + 2 \sum_{0 < \mu_m < \omega}^{m \leq M} a(\mu_m) \leq \frac{U}{1 - \omega/\Omega}.$$

As $N \to \infty$, we have $M \to \infty$, and

$$a_0 + 2 \sum_{0 < \mu_m < \omega} a(\mu_m) \leq \frac{U}{1 - \omega/\Omega}. \tag{30}$$

This gives us the inequality (23). Essentially this is Fekete's proof.

6. We now consider odd functions $\psi(-x) = -\psi(x)$, and the corresponding generalized sine-series. We assume again that

$$\sup_x \frac{1}{x} \int_0^x \psi^2(t)\,dt < \infty$$

and that

$$\lim_{x\to\infty} \frac{1}{x} \int_0^x \psi(t) \sin \lambda t \, dt = b(\lambda) = M\{\omega(t) \sin \lambda t\}$$

exists for all $\lambda > 0$. Then, again, $b(\lambda)$ vanishes, except at an enumerable set of λ-values. We denote in a certain order by $\lambda_1, \lambda_2, \cdots$ those λ for which $b(\lambda) \neq 0$, and call the series

$$\sum_1^\infty b_n \sin \lambda_n x \sim \psi(x) \quad \text{where} \quad b_n = 2b(\lambda_n), \tag{32}$$

the Fourier expansion of $\psi(x)$. We denote by $\rho_1, \rho_2, \cdots$ a subsequence of $\{\lambda_n\}$, consisting of linearly independent numbers, and by $\{\mu_n\}$ the subsequence of those λ_n, which can be represented in the form

$$\lambda_n = \sum_{\nu=1}^{h} r_\nu \rho_\nu, \quad h = h(n), \quad \text{where} \quad r_\nu = r_\nu(n) \quad \text{are rational.}$$

Again $0 < \omega < \Omega$, and we suppose

$$b(\mu_n) \geq 0 \quad \text{for} \quad 0 < \mu_n < \Omega.$$

If we associate with (32) the polynomials

$$T_q(x) = M\left\{ \psi(t) \sum_{\nu_1=-P}^{P} \cdots \sum_{\nu_q=-P}^{P} \left(1 - \frac{|\nu_1|}{P}\right) \cdots \left(1 - \frac{|\nu_q|}{P}\right) \sin t \frac{\nu_1\rho_1 + \cdots + \nu_q\rho_q}{Q} \sin x \frac{\nu_1\rho_1 + \cdots + \nu_q\rho_q}{Q} \right\}, \quad q \geq 1,$$

then by virtue of (31) we can write

$$T_q(x) = \sum k_n^{(q)} b_n \sin \lambda_n x,$$

where

$$k_n^{(q)} = \prod_{a=1}^{q} \left(1 - \frac{|\nu_a|}{P}\right) \text{ whenever } 0 < \lambda_n = \frac{1}{Q} \sum_{a=1}^{q} \nu_a \rho_a$$

and $k_n^{(q)} = 0$ for all other cases. We now have

$$T_q(x) = M\{\psi(t) \tfrac{1}{2}[K_q(t - x) + K_q(t + x)]\},$$

and on assuming

$$|\psi(t)| \leq U, \qquad t > 0$$

we get

$$|T_q(x)| \leq U, \qquad x > 0. \tag{33}$$

Again, introducing the polynomials

$$S_\mu(x, q) = \sum_{0<\lambda_n\leq\omega} \left(1 - \frac{\lambda_n}{\mu}\right) k_n^{(q)} b_n \sin \lambda_n x, \qquad \mu < \Omega,$$

and applying formulas (29) we obtain

$$S_\mu(x, q) = \frac{2}{\pi}\int_0^\infty \tfrac{1}{2}[T_q(x+2t) - T_q(x-2t)] \frac{\sin^2 \mu t}{\mu t^2}\, dt.$$

From this and (33) follows

$$(34) \qquad | S_\mu(x, q) | \leq U, \qquad x > 0.$$

At this point we simplify Fekete's argument, replacing a theorem of S. Bernstein by the following

LEMMA. *Given a generalized sine-polynomial*

$$S(x) = \sum_1^n b_\nu \sin \lambda_\nu x, \qquad \lambda_\nu > 0, \qquad (\nu = 1, 2, \cdots, n),$$

assume

$$b_\nu > 0, \ (\nu = 1, 2, \cdots, n), \ \textit{and} \ S(x) \leq U \ \textit{for} \ 0 < x < \xi.$$

Then

$$\sum_{\lambda_\nu\leq\omega} \lambda_\nu b_\nu \leq \frac{\alpha}{\sin^2 \alpha}\, \omega U < 1.38\omega U, \ \text{for} \ \omega \geq \frac{2\alpha}{\xi},$$

where α is defined in Theorem 2.

For the proof consider

$$\int_0^x S(t)\, dt = \sum_1^n b_\nu \frac{1-\cos \lambda_\nu x}{\lambda_\nu} = \tfrac{1}{2}x^2 \sum_1^n \lambda_\nu b_\nu \left(\frac{\sin \frac{1}{2}\lambda_\nu x}{\frac{1}{2}\lambda_\nu x}\right)^2,$$

from which we find

$$\tfrac{1}{2}x^2 \sum_{\lambda_\nu\leq\omega} \lambda_\nu b_\nu \left(\frac{\sin \frac{1}{2}\lambda_\nu x}{\frac{1}{2}\lambda_\nu x}\right)^2 \leq \int_0^x S(t)\, dt \leq xU.$$

Since $\omega x \leq \pi$,

$$\frac{\sin \frac{1}{2}\lambda_\nu x}{\frac{1}{2}\lambda_\nu x} \geq \frac{\sin \frac{1}{2}\omega x}{\frac{1}{2}\omega x}, \qquad \lambda_\nu \leq \omega,$$

and

$$\left(\frac{\sin \frac{1}{2}\omega x}{\frac{1}{2}\omega x}\right)^2 \frac{x}{2} \sum_{\lambda_\nu\leq\omega} \lambda_\nu b_\nu \leq U.$$

Finally, choosing $x = 2\alpha/\omega < \pi/\omega$, we have

$$\sum_{\lambda_\nu\leq\omega} \lambda_\nu b_\nu \leq \frac{\alpha}{\sin^2 \alpha}\, \omega U, \ \text{for} \ \frac{2\alpha}{\omega} \leq \xi.$$

Applying this lemma to the polynomial $S_\mu(x, q)$, we get, from (34),

$$(35) \qquad \sum_{0<\lambda_n\leq\omega}\left(1-\frac{\lambda_n}{\mu}\right)k_n{}^{(q)}\lambda_n b_n \leq \frac{\alpha}{\sin^2\alpha}\,\omega U, \qquad 0<\omega<\Omega,$$

and passing to the limit $q\to\infty$,

$$2\sum_{0<\mu_m\leq\omega}\left(1-\frac{\mu_m}{\mu}\right)\mu_m b(\mu_m) \leq \frac{\alpha}{\sin^2\alpha}\,\omega U.$$

But $1-\mu_m/\mu \geq 1-\omega/\mu$, hence

$$2\sum_{0<\mu_m\leq\omega}\mu_m b(\mu_m) \leq \frac{\omega U\dfrac{\alpha}{\sin^2\alpha}}{1-\dfrac{\omega}{\mu}};$$

and, as $\mu\to\Omega$, this takes the form

$$(36) \qquad 2\sum_{0<\mu_m\leq\omega}\mu_m b(\mu_m) \leq \frac{\omega U\dfrac{\alpha}{\sin^2\alpha}}{1-\dfrac{\omega}{\Omega}},$$

which is sharper than Fekete's inequality

$$2\sum_{0<\mu_m\leq\omega}\mu_m b(\mu_m) \leq \frac{\Omega U}{1-\dfrac{\omega}{\Omega}},$$

if we restrict ourselves to $\omega\,\dfrac{\alpha}{\sin^2\alpha}<\Omega$.

From (36) it follows that

$$\Big|\sum_{0<\lambda_n\leq\omega}k_n{}^{(q)}\frac{\lambda_n}{\omega}b_n\sin\lambda_n x\Big| \leq \frac{U\dfrac{\alpha}{\sin^2\alpha}}{1-\dfrac{\omega}{\Omega}},$$

and from (34), for $\mu=\omega$,

$$\Big|\sum_{0<\lambda_n\leq\omega}k_n{}^{(q)}\left(1-\frac{\lambda_n}{\omega}\right)b_n\sin\lambda_n x\Big| \leq U.$$

Thus, writing $p_m{}^{(q)}$ for $k_n{}^{(q)}$ if $\lambda_n=\mu_m$,

$$2\Big|\sum_{0<\mu_m\leq\omega}p_m{}^{(q)}b(\mu_m)\sin\mu_m x\Big| \leq U\left(1+\frac{\alpha}{\sin^2\alpha\left(1-\dfrac{\omega}{\Omega}\right)}\right).$$

From (36) follows, as Fekete proved, that $\sum_{0<\mu_m\leq\omega}b(\mu_m)\sin\mu_m x$ converges absolutely, hence by (27)

$$(37) \qquad \lim_{q\to\infty}2\sum_{0<\mu_m\leq\omega}p_m{}^{(q)}b(\mu_m)\sin\mu_m x = 2\sum_{0<\mu_m\leq\omega}b(\mu_m)\sin\mu_m x=\psi_\omega(x),$$

$$(38) \qquad |\psi_\omega(x)| \leq U\left(1+\frac{\alpha}{\sin^2\alpha\left(1-\dfrac{\omega}{\Omega}\right)}\right).$$

If $\{\rho_k\}$ is a basic sequence of $\{\lambda_k\}$, it follows that

$$\psi_\omega(x) = \sum_{0<\lambda_n\leq\omega} b_n \sin \lambda_n x, \qquad 0 < \omega < \Omega,$$

and this series converges absolutely. If, in particular, all $a_n \geq 0$ and all $b_n \geq 0$, then Ω is arbitrary, and, letting $\Omega \to \infty$, from (22) and (30) we find

$$\sum_{0<\lambda_n\leq\omega} a_n \leq U, \qquad |\psi_\omega(x)| \leq U\left[1 + \frac{\alpha}{\sin^2 \alpha}\right].$$

Hence for any real function $f(x)$ which satisfies

$$|f(x)| \leq U, \qquad -\infty < x < +\infty$$
$$M\{f(t)e^{-i\lambda t}\} = c(\lambda); \qquad c(\lambda_n) = \tfrac{1}{2}(a_n - ib_n),$$

we have

$$|a_0 + \sum_{0<\lambda_n\leq\omega} (a_n \cos nx + b_n \sin nx)| \leq U\left[2 + \frac{\alpha}{\sin^2 \alpha}\right].$$

The fact that the functions $\phi_\omega(x)$, $\psi_\omega(x)$ are u. a. p. and that their Fourier expansions are (17) and (29) respectively follows in exactly the same way as in Fekete's theorem.

Analogous argument can be applied to derive corresponding theorems for trigonometric integrals (cf. Szász [5]) and also to improve upon some results concerning Fourier series with $na_n \geq -K$, $nb_n \geq -K$ and more general classes (cf. Szász [4]).

UNIVERSITY OF CINCINNATI.

REFERENCES.

[1] L. Fejér, "On a theorem of Paley," *Bulletin of the American Mathematical Society*, vol. 40 (1934), pp. 469-475.

[2] M. Fekete, "On generalized Fourier series with non-negative coefficients," *Proceedings of the London Mathematical Society*, ser. 2, vol. 29 (1935), pp. 321-333.

[3] R. E. A. C. Paley, "On Fourier series with positive coefficients," *Journal of the London Mathematical Society*, vol. 7 (1932), pp. 138-144.

[4] O. Szász, "Convergence properties of Fourier series," *Transactions of the American Mathematical Society*, vol. 37 (1935), pp. 483-500.

[5] ———, "Über die Partialsummen Fourierscher Reihen," *Mathematischer und Naturwissenschaftlicher Anzeiger der Ungarischen Akademie der Wissenschaften*, 1937. In print.

ON DIRICHLET'S PROBLEM.*

By C. Carathéodory.

During the last 75 years Dirichlet's Problem has attracted the attention of many of the best scholars of the time with the result that this problem, which once was considered as one of the most difficult subjects in higher mathematics, can be treated now in a very elementary way.

Carl Neumann and *H. A. Schwarz* were the first to consider the problem of Dirichlet with modern methods, but while the road laid out by *C. Neumann* leads to the theory of integral equations, it is the work of *Schwarz* which has influenced most of the subsequent proofs for Dirichlet's problem. It contains already the fundamental idea of building up by a convergent iterative process very general classes of solutions by using the particular solutions given by Poisson's integral for the sphere.

The "sweeping out" process devised fifteen or twenty years later by *Poincaré* lies in the same line of thought but represents an immense advance if one compares it with the early method which *Schwarz* had developed chiefly for the case of the logarithmic potential and for very special boundaries. The splendid way in which *Poincaré*, taking advantage of the progresses made in the meantime by the theory of point sets, deals with the problem, has given many hints to younger scientists and has contributed to many advances made subsequently in different fields of mathematical science. But originally the method of *Poincaré* was very complicated and cumbersome; he felt, in fact, obliged to split the functions to which he was applying his process into two parts for which his convergence proof was holding, and to do this he had to approximate these functions by polynomials. Thus many elements foreign to potential theory were implied in his proof and the feeling that this proof was likely to be simplified and made clearer in its outline was present from the beginning to many students of the topic.

But it was not until 1912 that a most decisive and very surprising turn was given to the whole subject in a short note of 3 pages by *H. Lebesgue.* In this note it was shown that Poisson's Integral, which up to that time had been the primordial tool with which the convergent processes were built up, could be discarded altogether and replaced by much weaker operators connected with the property of the mean which characterizes harmonic functions. These

* Received February 28, 1937.

results of *Lebesgue*, though widely quoted, are not as universally known as would have been the case if their importance had been fully appreciated from the outset.

The great variety of convergent processes outlined above has induced me to look at the common root out of which all these processes originate. The result is contained in the second chapter of this memoir. It consists of a theorem whose assumptions are taken general enough so as to disclose the very reasons for which all the different methods used previously are equally successful.

In the foregoing chapter I have tried to give a very elementary treatment of the principal properties of harmonic functions culminating in the existence proof for Dirichlet's problem devised by *O. Perron* and very much simplified by *T. Radò* and *F. Riesz.* I have done this in order to show how the whole theory can be condensed if one puts systematically from the outset Poisson's Integral in the limelight.

Chapter I. Existence of the Solution of Dirichlet's Problem.

1. Definition of harmonic functions. Functions $U(x_1, \cdots, x_n)$, which are defined in some region R of the Euclidean space with the coördinates

$$x_1, \cdots, x_i, \cdots x_n$$

are called harmonic in that region if they are first of all continuous, if secondly considering them as functions of every single variable x_i alone they have continuous first derivatives $\partial U/\partial x_i$ and finite second derivatives $\partial^2 U/\partial x_i^2$ and if the expression

$$\Delta U = \frac{\partial^2 U}{\partial x_1^2} + \cdots + \frac{\partial^2 U}{\partial x_n^2} \tag{1.1}$$

formed with these second derivatives vanishes identically throughout R.

We shall see later on (§ 6) that every harmonic function is analytic and that consequently all of the successive partial derivatives exist and are continuous in all the variables. But to prove this we do not need to assume the existence of other partial derivatives than those which we have mentioned above or to assume that these derivatives are bounded or that the first derivatives $\partial U/\partial x_i$ are continuous functions of other variables than that which appears in the denominator.

In order to study under these assumptions the most general solutions of the equation

$$\Delta U = 0 \tag{1.2}$$

we must first obtain some special solutions of that equation.

Taking a point $x_i{}^0$ in our space and putting

$$r^2 = (x_1 - x_1{}^0)^2 + \cdots + (x_n - x_n{}^0)^2 \tag{1.3}$$

we have for every function of the form $F(r)$ the identity

$$\Delta F = F''(r) + F'(r)\,\frac{n-1}{r}. \tag{1.4}$$

The most general harmonic functions of the form $F(r)$ are therefore given by the equation

$$F(r) = \frac{A}{r^{n-2}} + B \tag{1.5}$$

for $n \geqq 3$ and by

$$F(r) = A \log r + B \tag{1.6}$$

for $n = 2$.

2. Poisson's integral. Taking first $n \geqq 3$ we consider a sphere of radius a and on some radius of this sphere two points at the distances $\lambda\,(\lambda < a)$ and a^2/λ from the center respectively. It follows then easily by the consideration of similar triangles that for points Q taken anywhere on the surface of the sphere one has

$$\frac{r'_0}{r_0} = \frac{a}{\lambda}$$

and herefrom that the function

$$V_\lambda = \sigma\left[\left(\frac{1}{r}\right)^{n-2} - \left(\frac{a}{\lambda r'}\right)^{n-2}\right] \tag{2.1}$$

is a harmonic function which vanishes identically at the surface of the sphere. We have now from the triangles of the figure on page 712

$$\begin{aligned}\frac{\lambda^2}{a^2}\,r'^2 &= \frac{\lambda^2}{a^2}\left[\frac{a^4}{\lambda^2} + \rho^2 - 2\,\frac{a^2}{\lambda}\,\rho\cos\theta\right]\\ &= a^2 + \frac{\lambda^2}{a^2}\,\rho^2 - 2\lambda\rho\cos\theta,\\ r^2 &= \lambda^2 + \rho^2 - 2\lambda\rho\cos\theta,\end{aligned}$$

so that we can write

$$\frac{\lambda^2}{a^2}\,r'^2 = r^2 + u \tag{2.2}$$

with

$$u = \frac{(a^2 - \lambda^2)(a^2 - \rho^2)}{a^2}. \tag{2.3}$$

By these formulas we can write

$$\left(\frac{a}{\lambda r'}\right)^{n-2} = (r^2 + u)^{-[(n-2)/2]}; \tag{2.4}$$

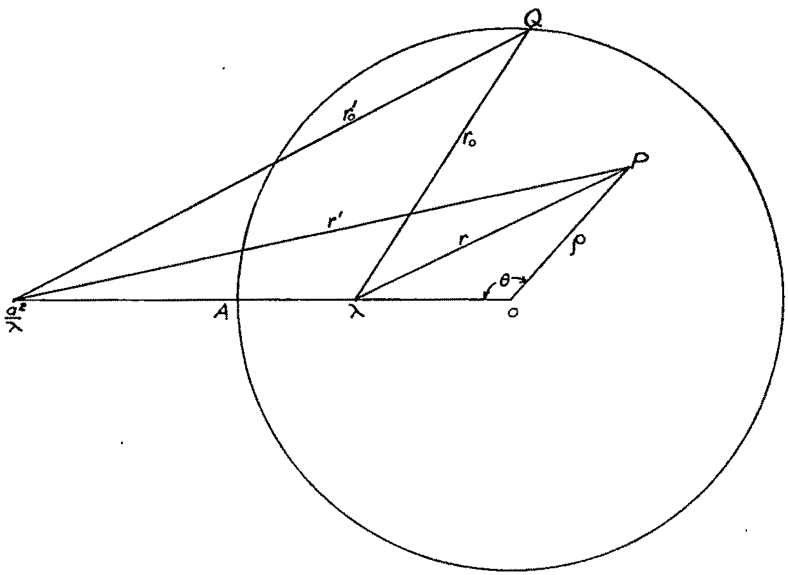
therefore if we put

$$\omega = a^2 - \lambda^2 \tag{2.5}$$

and, observing that r depends not only upon the x_i but also upon λ,

$$\left(r^2 + \frac{(a^2 - \rho^2)\,\omega}{a^2}\right)^{-[(n-2)/2]} = \phi(x_i, \lambda, \omega), \tag{2.6}$$

we can calculate the right hand of (2. 4) by $\phi(x_i, \sqrt{a^2 - \omega}, \omega)$ for the value of ω given by (2. 5). With these notations we can therefore replace (2. 1) by



$$V_\lambda = -\sigma\omega \frac{\phi(x_i, \sqrt{a^2 - \omega}, \omega) - \phi(x_i, \sqrt{a^2 - \omega}, 0)}{\omega}. \tag{2.7}$$

We shall now take $\sigma\omega$ to be a constant for varying ω and make use of the fact that the function

$$\psi(x_i, \lambda, \omega) = \frac{\phi(x_i, \lambda, \omega) - \phi(x_i, \lambda, 0)}{\omega} \tag{2.8}$$

is analytic if the x_i vary in a suitable domain and that, $|a - \lambda|$ and $|\omega|$ being sufficiently small, the point $\omega = 0$ is not excluded from the domain of regularity of that function. We have in fact

$$\psi(x_i, \lambda, 0) = \frac{\partial\phi}{\partial\omega}\bigg|_{\omega = 0} = -\frac{n-2}{2}\,\frac{1}{r^n}\,\frac{a^2 - \rho^2}{a^2}. \tag{2.9}$$

Taking therefore

$$\sigma\omega = C\frac{2a^2}{n-2} \tag{2.10}$$

the function V_λ takes at the limit $\omega = 0$ or, which is the same, at the limit $\lambda = a$ the value

$$V_a = \sigma\omega\psi(x_i, a, 0) = C\frac{a^2-\rho^2}{r^n}. \tag{2.11}$$

Now we have for real and positive values of ω

$$\Delta\psi(x_i, \sqrt{a^2-\omega}, \omega) = -\frac{n-2}{2a^2C}\Delta V_\lambda = 0$$

and the analytic function $\Delta\psi(x_i, \sqrt{a^2-\omega}, \omega)$ being regular for $\omega = 0$ we see that the right hand of (2.11) is *harmonic* at every real point of the space except at the point A of the above figure.

3. In order to show that the formula (2.11) holds also for the case of the two dimensional plane we observe that the function

$$V_\lambda = \sigma\left[\log\frac{\lambda}{a}r' - \log r\right] \tag{3.1}$$

is harmonic and vanishes at the boundary of the circle with radius a and that by using our previous notations we can write

$$\begin{aligned} V_\lambda &= \frac{\sigma}{2}\left[\log(r^2+u) - \log r^2\right] \\ &= \frac{\sigma}{2}\log\left(1+\frac{a^2-\rho^2}{a^2r^2}\omega\right) \end{aligned} \tag{3.2}$$

Putting now

$$\phi(x_i, \lambda, \omega) = \log\left(1+\frac{a^2-\rho^2}{a^2r^2}\omega\right) \tag{3.3}$$

we find that (2.7) holds also in the case $n = 2$ and that we get by the same method as above the relation (2.11), if we choose σ to satisfy the relation

$$\sigma\omega = C\cdot 2a^2. \tag{3.4}$$

4. We consider now a continuous function $V(Q)$ on the boundary of the sphere (or the circle) of radius a and form the integral

$$V(P) = C\int V(Q)\,\frac{a^2-\rho^2}{r^n}\,do \tag{4.1}$$

in which P is a point of the interior of the considered domain at the distance ρ from its center and r designates the distance between P and Q. One verifies

by differentiation that the function $V(P)$ is harmonic in the interior of the sphere. It is easy to calculate this function for the case when $V(Q)$ is a constant. In fact $V(P)$ is then a harmonic function of the form $F(\rho)$ which is bounded in the neighborhood of the center of the sphere and which by the formulas (1. 5) and (1. 6) must, therefore, in any case be a *constant*. At the center of the sphere we have $\rho = 0$ and $r \equiv a$ and therefore

$$V(P) = V(Q)\, \frac{C}{a^{n-2}} \int do.$$

This last integral has the value $\alpha_n a^{n-1}$ where α_n is a known constant depending only on the dimension n of the space. In order to have $V(P) \equiv V(Q)$ we must therefore take

$$C = \frac{1}{\alpha_n a} \tag{4. 3}$$

and get the final formula

$$V(P) = \frac{1}{\alpha_n a} \int V(Q)\, \frac{a^2 - \rho^2}{r^n}\, do \tag{4. 4}$$

which is known by the name of *Poisson's integral.*

5. From the last reasonings we infer that the identity holds

$$\frac{1}{\alpha_n a} \int \frac{a^2 - \rho^2}{r^n}\, do \equiv 1. \tag{5. 1}$$

With the aid of this identity *H. A. Schwarz* has proved that the function $V(P)$ represented by the relation (4. 4) is continuous on the *closed* sphere and that the boundary values of $V(P)$ coincide with the given values of $V(Q)$.

We omit the proof of this theorem of *Schwarz* which is classical and is contained in all the books of reference dealing with our subject.

6. The fundamental theorem. The proposition from which we can deduce nearly all the principal properties of harmonic functions consists in the statement that in the neighborhood of a point we may always represent any given harmonic function by a suitably chosen Poisson's integral.

We consider a function $f(P)$ which is continuous on a closed sphere and harmonic in the interior of it according to the definition of § 1, and calculate the function $V(P)$ which is given by the integral of Poisson and which possesses the same boundary values as $f(P)$. Furthermore we calculate by the integral of Poisson a function $\psi(P)$ which is harmonic in the interior of the sphere and takes on its surface the same boundary values as the function

$$\tfrac{1}{2}x_1^2.$$

Putting now

$$\Phi(P) = (f(P) - V(P)) + \alpha(\psi(P) - \tfrac{1}{2}x_1^2) \tag{6.1}$$

we observe that the function $\Phi(P)$ is continuous on the closed sphere, vanishes at its boundary and that the equality

$$\Delta\Phi + \alpha = 0 \tag{6.2}$$

holds at all the interior points of the sphere. From this we infer that for positives values of the parameter α

$$\Phi \geqq 0 \tag{6.3}$$

everywhere in the sphere. For if at some point of the sphere Φ were negative there should exist at least one point P_0 of the interior of the sphere at which $\Phi(P)$ would attain its minimum value. At such a point we should have necessarily

$$\frac{\partial\Phi}{\partial x_i} = 0, \qquad \frac{\partial^2\Phi}{\partial x_i^2} \geqq 0 \qquad (i = 1, 2, \cdots, n), \tag{6.4}$$

i. e., conditions which contradict (6. 2). Letting now α tend towards zero we find that we must have everywhere in the sphere $V(P) \leqq f(P)$. We prove in the same way the inequality $-V(P) \leqq -f(P)$ so we get finally

$$V(P) \equiv f(P) \tag{6.5}$$

and our fundamental theorem is thus proved.

7. The theorem of Gauss and its converse. In order to calculate Poisson's integral at the center O of the sphere we have to put in (4. 4)

$$\rho = 0, \qquad r \equiv a$$

and we get

$$V(O) = \frac{1}{\alpha_n a^{n-1}} \int V(Q)\, do. \tag{7.1}$$

This formula shows that the value of $V(P)$ at the center O of our sphere is equal to the mean of the values $V(Q)$ at the surface of that figure.

Considering the fundamental theorem of the previous section we see that this property applies to the most general harmonic functions and to every sphere which lies entirely inside of the region for which the function is defined. This statement is called the theorem of Gauss.

8. Take now a continuous function $g(P)$ for which the theorem of Gauss holds for every sphere lying inside of a region R. Suppose that $g(P)$ is not a

constant and call M and m the upper and the lower bound of $g(P)$ when P is describing the region. We are going to show that at every point P_0 of R we always have

$$m < g(P_0) < M \tag{8.1}$$

the equality signs being excluded. Call R' that part of R at each point Q of which we have

$$g(Q) \neq g(P_0). \tag{8.2}$$

The point set R' is an open set which ($g(P)$ not being constant) is not void and does not fill out the region R. There exist therefore at least one boundary point P_1 of R' lying inside of R and at this point we have

$$g(P_1) = g(P_0). \tag{8.3}$$

Furthermore there exists within R at least a sphere with center P_1 on whose surface there is at least one point Q of R'. The function $g(P)$ cannot be constant throughout the surface of this sphere for in that case we should have by Gauss' theorem $g(P_1) = g(Q)$ and this is contrary to our hypothesis. The function $g(P)$ being continuous and not constant on the surface of the sphere the mean $g(P_1)$ of these values must actually lie between M and m and by (8.3) this is equivalent to the relation (8.1) we wanted to prove.

9. From the last result it follows that if a function $g(P)$ is continuous on a closed region $\bar{R}$ and vanishes at every point of the boundary of R, if furthermore Gauss' theorem holds for every sphere contained in R the function $g(P)$ vanishes identically. If indeed $g(P)$ were not a constant there would be points of R at which say $g(P) > 0$ and then it would exist at least one point P_0 in R at which $g(P)$ would attain its maximum value. But this contradicts the result of the last section.

10. It follows herefrom that every function $f(P)$ for which the theorem of Gauss holds for every sphere which lies inside of a region R must be harmonic. Consider indeed the harmonic function $V(P)$ which has the same values as $f(P)$ on the surface of such a sphere. The conditions stated in § 9 apply to the function

$$g(P) = f(P) - V(P)$$

in the sphere in which it is defined. We therefore have

$$f(P) \equiv V(P)$$

which proves that $f(P)$ is harmonic.

We have so far obtained two very remarkable results: in first line we have seen that harmonic functions may be as well characterized by the property expressed by Gauss' theorem as by the differential equation $\Delta f = 0$. It is very surprising at first sight that the property of the mean involves the functions for which it holds to be analytic.

Secondly we see in using the result of § 9 that there is *at most* one function which being harmonic inside of a given region is continuous at the boundary and which has there prescribed values. Dirichlet's problem has therefore either no solution at all or just one single solution. For the sphere there exists always a solution given by Poisson's integral. In order to make a choice between these possibilities for more general regions we must study with more detail the properties of harmonic functions.

11. Harnack's theorem. We suppose in the formula (4. 4) for Poisson's integral the function $V(Q)$ not to be negative. As we have at the center O of the sphere

(11. 1) $$V(O) = \frac{1}{\alpha_n a^{n-1}} \int V(Q)\, do$$

and as for any point P interior to a concentric sphere of radius $a\theta$ (the number θ being positive and less than one) we have

(11. 2) $$\rho \leqq a\theta, \quad (1-\theta)a \leqq r \leqq (1+\theta)a$$

uniformly for every point Q of the surface, we can write

(11. 3) $$V(O)\frac{1-\theta^2}{(1+\theta)^n} \leqq V(P) \leqq V(O)\frac{1}{(1-\theta)^n}.$$

Suppose now that at a point P' of the smaller sphere the inequality

(11. 4) $$V(P') < M$$

holds. We then have by (11. 3)

(11. 5) $$V(O) < \frac{(1+\theta)^{n-1}}{(1-\theta)} M$$

and consequently for every point P'' of the same sphere

(11. 6) $$V(P'') < \frac{(1+\theta)^{n-1}}{(1-\theta)^{n+1}} M.$$

Furthermore we have always

(11. 7) $$| V(P'') - V(P') | \leqq V(0)\frac{(1+\theta)^{n-1} - (1-\theta)^{n+1}}{(1-\theta)(1-\theta^2)^{n-1}}.$$

CALCUTTA UNIVERSITY LIBRARY
Sec. No. .... Book No. ....

12. We consider now the family of functions $U(P)$ which being harmonic inside of a region are at every point of R not smaller than a given constant α and besides at a certain point P_0 of the same region not greater than a second constant β. If we apply the inequality (11. 6) to the functions $(U(P) - \alpha)$ we can easily show in using chains of spheres overlapping one another that the functions of our family have a common finite upper bound on any given closed set of points lying inside of R.

This being the case the inequality (11. 7) shows that the limiting oscillation of our family of functions vanishes at every interior point of R and that the family is therefore a normal family in the sense of *Montel*.

Consequently we can extract from every sequence U_j of functions of our family a subsequence U_{j_k} which converges continuously towards a function $V(P)$ at every point of R. The convergence must therefore be uniform on every closed subset of R and we infer herefrom that for the limiting function $V(P)$ the theorem of Gauss holds for every sphere contained in R and that $V(P)$ must therefore itself be harmonic.

We can state this result more briefly by saying that our family is not only *normal* but also *closed*; i. e. that every limiting function of functions of our family belongs itself to the same family.

The well known theorem of *Harnack* is very closely connected with these last results which are but slightly more general (though much more convenient for the applications than the older theorem).

13. Superharmonic and Subharmonic functions. Classes of functions which have been named in recent times super- and subharmonic have been used rather early in the literature. *Poincaré* has made use of them; but it was not until they were carefully studied by *F. Riesz* that their great importance for Potential theory has appeared. *Riesz* has made use of such functions under the assumption that they are semicontinuous. For our purposes it will be sufficient to use the following definition: a function $S(P)$ is called superharmonic in a region R if being continuous in that region $S(P)$ is never smaller than the mean of $S(Q)$ taken for the surfaces of every sphere of center P lying entirely in R and whose radii are less than a positive number $r_0(P)$ whose magnitude varies with P. The definition of subharmonic functions $s(P)$ is the same but for the sign of the inequality.

Calling m the lower bound of $S(P)$ in the region R one shows that if $S(P)$ be superharmonic and not a constant we have always

$$m < S(P). \tag{13.1}$$

Taking an arbitrary point P_0 in R we construct just as in § 8 a sphere with center P_1 lying in R and possessing at its surface at least one point Q for which

$$S(Q) \neq S(P_1) \tag{13.2}$$

the center P_1 of the sphere having the further property that $S(P_1) = S(P_0)$. If $S(P)$ is not constant on the surface of this sphere the mean of the function taken over that sphere is actually greater than m and $S(P)$ being superharmonic we have $S(P_1) > m$. But if $S(P)$ is constant throughout the surface of the sphere we have by the condition of superharmonicity $S(P_1) \geqq S(Q)$ and by (13.2)

$$S(P_1) > S(Q) \geqq m. \tag{13.3}$$

In both cases the relation (13.1) is verified at the point P_0.

Calling M the upper bound of a subharmonic function $s(P)$ we show in like manner that either $s(P)$ is a constant or that we have for every point of R

$$s(P) < M.$$

It follows that a superharmonic function $S(P)$ which is continuous on a *closed* region $\bar{R}$ and which is not negative at the boundary of R cannot have negative values in R and that under analogous conditions a subharmonic function $s(P)$ is never positive.

14. We shall make use of some properties of the superharmonic functions to which *F. Riesz* was the first to pay attention.

Consider first a sphere lying entirely in a region R in which a superharmonic function $S(P)$ is defined and call $V(P)$ the harmonic function (calculated by Poisson's Integral) which coincides with $S(P)$ at the boundary of that sphere. Applying the reasoning at the end of the previous section to the function $(S(P) - V(P))$ which is superharmonic inside the sphere and which vanishes at its surface we have

$$V(P) \leqq S(P). \tag{14.1}$$

15. We consider now a function $S^*(P)$ which is equal to $S(P)$ outside of the sphere which we have just considered and equal to $V(P)$ inside of this same sphere. As we have everywhere $S^*(P) \leqq S(P)$ the characteristic property for the function $S^*(P)$ to be superharmonic is verified outside and on the boundary of our sphere. But the same holds obviously also for the points inside of the sphere as $S^*(P)$ coincides with the harmonic function $V(P)$ at these points.

Thus $S^*(P)$ is superharmonic in the whole of the region R. This property of superharmonic functions which consists in the possibility to replace such a function by another which is nowhere greater and which is harmonic in a given sphere is of paramount importance for the sequel.

16. Finally we consider two superharmonic functions $S_1(P)$ and $S_2(P)$ defined in the same region and call $S(P)$ the function which at every point of R is equal to the smaller of both numbers $S_1(P)$ and $S_2(P)$. It is readily seen that $S(P)$ is also superharmonic in R. For at a point P_0 of R we have say

$$S(P_0) = S_1(P_0)$$

and by assumption $S_1(P_0)$ is not smaller than the mean of that function taken over the surface of a sufficient small sphere with center P_0. Then $S_1(P_0)$ or which is the same $S(P_0)$ cannot be smaller than the mean taken over the surface of the same sphere of the function $S(P)$ because we have everywhere $S(P) \leqq S_1(P)$.

17. Existence of the solution of Dirichlet's Problem. We consider on the boundary $(\bar{R} - R)$ of a given region R a continuous function $\phi(Q)$ and call according to *Perron* all the functions which are continuous in $\bar{R}$ and superharmonic in R and which at every point of the boundary are not less than $\phi(Q)$ an *upper function* $S(P)$ of our problem.

In a similar way a *lower function* $s(P)$ of the problem is a function which is continuous in $\bar{R}$ never greater than $\phi(Q)$ at the boundary of R and subharmonic in that region. According to § 13 the difference $(S(P) - s(P))$ of an upper and of a lower function being superharmonic in R and not negative at the boundary is nowhere negative in R.

If we call m and M the minimum and the maximum of $\phi(Q)$ on the boundary of R the constant function equal to M is an upper function of the problem and the constant function equal to m a lower function.

Accordingly we have at every point of R and for every upper function $S(P)$

$$S(P) \geqq m \tag{17.1}$$

and the lower bound $U(P)$ of all possible upper functions at a point P of R satisfies the conditions

$$m \leqq U(P) \leqq M. \tag{17.2}$$

We will show that $U(P)$ is harmonic in R.

18. We consider a sequence of points

$$P_1, P_2, \cdots, P_j, \cdots \tag{18.1}$$

which is everywhere dense in R and for each of these points say for P_j a sequence of upper functions

$$S_{j1}, S_{j2}, \cdots, S_{jk}, \cdots \tag{18.2}$$

which satisfy the condition

$$U(P_j) = \lim_{k=\infty} S_{jk}(P_j). \tag{18.3}$$

We call now $S_k(P)$ the function which at every point of R is equal to the smallest of the numbers

$$M, S_{1k}(P), S_{2k}(P), \cdots, S_{kk}(P).$$

According to § 17 this function is a bounded upper function and we have

$$U(P_j) = \lim_{k=\infty} S_k(P_j) \qquad (j = 1, 2, \cdots). \tag{18.4}$$

We replace now the functions S_k by the functions S^*_k which we have constructed in § 15 after having taken an arbitrary sphere σ inside of the region R.

The functions S^*_k are again upper functions for the problem for which relations analogous to (18.4) hold because we have at every point of R

$$U(P) \leqq S^*_k(P) \leqq S_k(P). \tag{18.5}$$

Now by § 12 the functions $S^*_k(P)$ belong to a normal family of harmonic functions inside of the sphere σ and form a sequence which is convergent at an everywhere dense set of points in this sphere. They must therefore converge towards a harmonic function $V(P)$ and we are going to show that one must have $V(P) = U(P)$ at *all* the points of the sphere σ. This is true by construction for all those points of the countable set (18.1) which lie inside σ. We take at random another point of σ which we shall call P_0 and we add it to the points of the set (18.1). We can then construct by the same process as above a new function $V'(P)$, which being harmonic inside of σ is equal to $U'(P)$ at every point of the set (18.1) and besides at the point P_0. The functions $V(P)$ and $V'(P)$ are both continuous in σ and they are equal to one another at the points of the set (18.1) which is everywhere dense in that sphere. They must therefore coincide in σ and we have in particular $U(P_0) = V(P_0)$. This shows that $U(P)$ is harmonic in σ and as this sphere may be chosen at random the function $U(P)$ must be harmonic everywhere in R.

19. If the closed region $\bar{R}$ has the property that Dirichlet's problem is solvable for any given continuous boundary values, there exist functions $V(P; Q_0)$ of P continuous on the closed set $\bar{R}$ which vanish at a given point Q_0 of the boundary of R, which are positive at all the other boundary points of R and harmonic in this region.

This leads to the following definition: a boundary point Q_0 of a given region (for which one does not know if Dirichlet's problem is solvable or not) is called a *regular* point of the boundary if a function $V(P; Q_0)$ with the properties stated above exists.

For the case that Q_0 is a regular point of the boundary it is very easy to construct an upper function of our problem which at the point Q_0 has a value not greater than $\phi(Q_0) + \epsilon$, the positive number ϵ being chosen at random. Take a sphere $\sigma(Q_0; \epsilon)$ with center Q_0 and such that at all the points Q of the boundary of R lying inside of that sphere we have

$$\phi(Q) < \phi(Q_0) + \epsilon. \tag{19.1}$$

Call μ the necessarily positive minimum of the values which $V(P; Q_0)$ takes in $\bar{R}$ at points lying not inside $\sigma(Q_0; \epsilon)$. Then the function

$$S(P) = \phi(Q_0) + \epsilon + \frac{M - m}{\mu} V(P; Q_0)$$

is an upper function of our problem which possesses the stated property. In the same way a lower function of the problem which at the point Q_0 has a value not smaller than $\phi(Q_0) - \epsilon$ can be found.

It follows then that the harmonic function $U(P)$, which we have defined in the previous section is continuous at every regular point of the boundary of R and that it takes at such points the prescribed value $\phi(Q_0)$.

Thus for the solution of Dirichlet's problem it is not only necessary but it is also sufficient that all the points of the boundary of R be regular. It is well known that *Lebesgue* has given examples of regions—even of simply connected regions—for which not all the boundary points are regular. In the sequel we shall restrict ourselves to regions which possess a regular boundary.

Chapter II. Construction of the Solutions of Dirichlet's Problem.

20. Statement of the problem. We consider a closed bounded region $\bar{R}$ whose boundary points Q are all regular in the sense of § 19. We try to define sequences of operators

$$L_1 f, L_2 f, \cdots, L_j f, \cdots \tag{20.1}$$

which enable us to calculate by a limiting process the harmonic functions $U(P)$ whose existence we have proved in the two previous sections.

More precisely we shall take an arbitrary function $f_1(P)$ which is continuous on the closed region $\bar{R}$ and which at the boundary points Q of that region satisfies the equality

$$f_1(Q) = \phi(Q), \tag{20.1}$$

the last function being continuous on the boundary of R and taken at random. Then we shall construct a sequence of functions $f_k(P)$ given recurrently by the formulas

$$(20.2) \qquad f_2(P) = L_1 f_1(P), \cdots, f_{k+1}(P) = L_k f_k(P), \cdots.$$

We shall show that the operators $L_j f$ can be chosen in such a way that the sequence $f_k(P)$ converges uniformly towards the solution $U(P)$ of Dirichlet's problem which possesses the given boundary values $\phi(Q)$.

21. We shall suppose that the operators $L_j f$ are defined for all functions $f(P)$ which are finite and continuous in the interior of R no matter if these functions can be extended to a continuous function on the closed region $\bar{R}$ or not or if these functions are bounded in R or not.

We suppose furthermore that the result $L_j f$ of each of our operations applied to functions of the above class must be a function of the same class.

Our operators $L_j f$ shall have now the following properties:

a) they are linear i. e.

$$L_j(f+g) = L_j f + L_j g$$
$$L_j(-f) = -L_j f$$

b) they are definite i. e. if $f(P) \geqq 0$ throughout R then $L_j f \geqq 0$ throughout R.

c) they do not change the value of harmonic functions i. e. if $U(P)$ is harmonic everywhere in R we have

$$L_j U(P) \equiv U(P).$$

22. If we take the operators $L_j f$ satisfying the above conditions our problem is already greatly simplified. It will indeed be sufficient to restrict the $L_j f$ in such a way that for functions $f_1(P)$ whose boundary values vanish identically the functions $f_k(P)$ considered in § 20 converge uniformly towards zero. For in the general case we can subtract from $f_1(P)$ the harmonic function $U(P)$ which has the same boundary values $\phi(Q)$ and whose existence is guaranteed and put

$$g_1(P) = f_1(P) - U(P), \quad g_{k+1}(P) = L_k g_k(P) \qquad (k = 1, 2, \cdots).$$

We have then for every value of k

$$f_k(P) = g_k(P) + U(P)$$

and if we know that the $g_k(P)$ converge towards zero the functions $f_k(P)$ must converge towards $U(P)$ as desired.

23. Superharmonic cover. Before stating the further properties of the linear operators $L_j f$ which will insure the convergence of our process we must study closer the consequences of the properties a), b) and c) of our operators.

We take an arbitrary function $f(P)$ which is continuous in the closed region $\bar{R}$ and vanishes at all the points Q of the boundary of R. We consider functions $U(P)$ harmonic everywhere in R (but not necessarily defined on the boundary of R) which satisfy the condition

$$f(P) \leqq U(P). \tag{23.1}$$

The lower bound of all these functions $U(P)$ considered at the same point P of R defines a function $S(P)$ which we shall call the *superharmonic cover* of $f(P)$.

Take a point P_0 inside of R and consider a sequence $U_j(P)$ taken among the harmonic functions defined above and such that

$$S(P_0) = \lim_{j=\infty} U_j(P_0). \tag{23.2}$$

As $f(P)$ is a bounded function it follows from (23.1) and (23.2) and from Harnack's theorem that the functions $U_j(P)$ belong to a normal family of functions. There exist therefore subsequences of the $U_j(P)$ which converge towards a function $U_{P_0}(P)$ harmonic in R and for which the following properties hold

$$S(P) \leqq U_{P_0}(P), \qquad S(P_0) = U_{P_0}(P_0). \tag{23.3}$$

From $f(P) \leqq U_{P_0}(P)$ follows that $U_{P_0}(P) \geqq 0$, if we remind ourselves of the properties of harmonic functions stated in § 9 and of the assumption that $f(P)$ vanishes at the boundary of R. From the second relation (23.3) we infer that everywhere in R we have

$$S(P) \geqq 0. \tag{23.4}$$

It is furthermore easy to verify that the superharmonic cover $S(P)$ is a continuous function. Taking in fact a sequence of points P_j which converge towards P_0 we infer on using both relations (23.3) that

$$\overline{\lim_{j=\infty}}\, S(P_j) \leqq S(P_0). \tag{23.5}$$

We have now

$$S(P_j) = U_{P_j}(P_j), \qquad U_{P_j}(P_0) \geqq S(P_0) \tag{23.6}$$

and on using the fact that the functions $U_{P_j}(P)$ belong to a normal family we find easily

(23. 7) $$\varliminf_{j=\infty} S(P_j) \geqq S(P_0).$$

The continuity of $S(P)$ is thus proved by the inequalities (23. 5) and (23. 7).

Making now use of the assumption that any point Q_0 of the boundary of R is a regular point we can construct functions $U(P)$ which are continuous on the closed region $\bar{R}$, which are harmonic in R and not less than the given function $f(P)$ and which at the point Q_0 have a value which is smaller than a given positive number ϵ (cf. § **19**). From this and from (23. 4) we infer that the superharmonic cover $S(P)$ is continuous at all the boundary points of R and that it vanishes at those points.

24. We observe now that we have

(24. 1) $$f(P) \leqq S(P) \leqq U_{P_0}(P)$$

and that consequently for every single operator

$$L_j f \leqq L_j S \leqq L_j U_{P_0}(P) = U_{P_0}(P).$$

It follows by the second relation (23. 3) that we must have at the point P_0

(24. 2) $$L_j f \leqq L_j S \leqq S.$$

Now we may choose P_0 anywhere in R and see that the inequalities (24. 2) hold in the whole region R.

We apply now these inequalities to the sequence of functions $f_k(P)$ defined by (20. 2). Calling S_k the superharmonic cover of f_k we have

$$f_{k+1}(P) = L_k f_k(P) \leqq S_k(P)$$

and it follows

(24. 3) $$S_{k+1}(P) \leqq S_k(P)$$

by the definition of the superharmonic cover of $f_{k+1}(P)$.

25. The monotonic sequence of functions

(25. 1) $$S_1(P) \geqq S_2(P) \geqq \cdots$$

converges towards a function

(25. 2) $$T(P) \geqq 0.$$

We are going to show that the convergence is continuous at every point P_0 of R. Taking a sequence of points P_k converging towards P_0 we have for any $k > j$

2

$$S_k(P_k) \leqq S_j(P_k)$$

and therefore

$$\overline{\lim_{k=\infty}} S_k(P_k) \leqq S_j(P_0).$$

As this holds for every value of j we have

$$\overline{\lim} S_k(P_k) \leqq T(P_0). \tag{25.3}$$

We take a subsequence $S_{k_j}(P)$ of our sequence such that

$$\underline{\lim} S_k(P_k) = \lim_{j=\infty} S_{k_j}(P_{k_j}). \tag{25.4}$$

We consider functions $V_j(P)$ harmonic in R for which the following conditions hold

$$V_j(P) \geqq S_{k_j}(P), \qquad V_j(P_{k_j}) = S_{k_j}(P_{k_j}). \tag{25.5}$$

The sequence of functions $V_j(P)$ belongs to a normal family; there exists a subsequence of these functions convergent continuously in R towards a function $V(P)$ and certainly

$$V(P) \geqq T(P).$$

We have therefore finally by (25.4)

$$\underline{\lim} S_k(P_k) = V(P_0) \geqq T(P_0) \tag{25.6}$$

and the comparison with (25.3) yields

$$\lim_{k=\infty} S_k(P_k) = T(P_0).$$

It follows by the general theory that the functions $S_k(P)$ converge uniformly towards $T(P)$ on any closed subset of R. If one considers that the non-negative function $S_1(P)$ is continuous on the closed region $\bar{R}$ and vanishes on the boundary of R and if one makes use of the fact that the sequence is monotonic and that the limiting function $T(P)$ is itself non-negative it follows that the sequence (25.1) converges uniformly towards $T(P)$ on the *closed* region $\bar{R}$.

26. The limiting function $T(P)$ is furthermore restricted if among the operators L_jf there are some which are *iterated*. We shall say that the operator L_qf is iterated if there exists an infinite sequence of increasing integers

$$q_1 < q_2 < \cdots \tag{26.1}$$

such that the operators L_{q_j} with the indices q_i be all of them identical with L_q

$$L_q f \equiv L_{q_j} f \qquad (j = 1, 2, \cdots). \tag{26.2}$$

Given a positive number ϵ; by the result at the end of the last section there are in the sequence (26.1) numbers q_j for which

$$S_{q_j} < T + \epsilon. \tag{26.3}$$

We have now by (24.2)

$$L_{q_j} f_{q_j} \leqq L_{q_j} S_{q_j} \leqq L_{q_j}(T + \epsilon) \tag{26.4}$$

and for the last number of this relation by (26.2)

$$L_{q_j}(T + \epsilon) = L_q T + \epsilon.$$

Calling $T'_q(P)$ the superharmonic cover of $L_q T(P)$ and remembering that S_{q_j+1} is the superharmonic cover of $L_{q_j} f_{q_j}$ we can therefore write

$$T(P) \leqq S_{q_j+1}(P) \leqq T'_q + \epsilon. \tag{26.5}$$

On the other hand we have for every value of the subscript k

$$L_q T \leqq L_q S_k \leqq S_k$$

and herefrom follows

$$T'_q(P) \leqq S_k(P)$$

and at the limit

$$T'_q(P) \leqq T(P). \tag{26.6}$$

Comparing with (26.5) we get finally

$$T'_q(P) \equiv T(P).$$

This result is not surprising: the operator $L_q f$ having been involved an infinite number of times already for the construction of the sequence of superharmonic functions $S_k(P)$ cannot modify any more the function $T(P)$ if we apply a construction similar to that which transforms S_q into S_{q+1}.

27. Certainly the conditions we have imposed up to now on our operators $L_j f$ are not sufficient to make the functions $f_k(P)$ converge towards zero. We could take the $L_j f$ all equal to the identical transformation and every single one of the conditions imposed so far upon them would be fulfilled.

But in that case a function $f_1(P)$ which is not identically zero would never yield a sequence $f_k(P)$ converging towards that number.

On the other hand if we can make sure that to every non-negative function $T(P) \neq 0$ corresponds at least one of the iterated operators $L_q f$ for which $T'_q(P) \neq T(P)$, by the result of the last section the function $T(P)$ must vanish identically and our sequence of functions $f_k(P)$ must then converge uniformly towards zero.

To prove this we consider the superharmonic covers of the functions $-f_k(P)$ and call them $-s_k(P)$. We have then

$$s_k(P) \leqq f_k(P) \leqq S_k(P) \tag{27.1}$$

and by our assumptions both of the sequences $s_k(P)$ and $S_k(P)$ must converge uniformly towards zero. The same is therefore true for the sequence of the $f_k(P)$.

28. A very general condition which we shall impose on our operators and which implies the just stated property is the following: we take a non-negative function $T(P)$ which is not a constant, which is continuous on the closed region R, which vanishes at the boundary of R and whose maximum value M is attained at an interior point P_0 of R. We assume furthermore that P_0 is such a point that there exists at least an $(n-1)$-dimensional plane α passing through P_0 and such that at every point Q of R lying on one side of α we have $T(Q) < M$. We require that under these assumptions for one at least of the iterated operators, say for $L_q f$, we have at the point P_0 itself

$$L_q T(P_0) < M. \tag{28.1}$$

We have only to prove that at a point P_0 which fulfills the above requirements we have always

$$T'_q(P_0) < M = T(P_0). \tag{28.2}$$

If there is only one point in the interior of R at which $T(P) = M$ we take this point for our P_0 and remark that by (28.1) the *convex cover* of $L_q T(P)$ lies underneath $T(P)$ at the point P_0 and that the superharmonic cover of $L_q T(P)$ lying underneath that just considered convex cover the inequality (28.2) is fulfilled.

If there be several points in R at which $T(P) = M$ these form a closed point set A lying inside of R and possessing no point on the boundary of R. We take two points P_1 and P_0 of this set whose distance from one another is

a maximum and consider the plane α passing through P_0 and normal to the vector $\overrightarrow{P_0P_1}$. All our conditions being fulfilled we must have

$$L_qT(P_0) < M \tag{28.3}$$

not only at P_0 itself but also inside a sphere σ with P_0 as center. The convex cover of $L_qT(P)$ is always less than M except perhaps on the smallest convex figure of the space which contains the points $(A - A\sigma)$. It is very easy to verify that the point P_0 is exterior to this figure and therefore we conclude just as before that (28.2) holds.

29. Our final result may be stated in the following way:

Take a bounded region R in n-space for which Dirichlet's problem has always a solution. Take a sequence of linear operators L_jf $(j = 1, 2, \cdots)$ defined for functions which are continuous in the interior of R, the result of every one of these operations being a function of the same kind. Suppose that the operators L_jf have the following properties:

1) *If $f(P) \geqq 0$ then $L_jf \geqq 0$ everywhere in R;*

2) *If $U(P)$ is harmonic everywhere in R then $L_jU \equiv U$;*

3) *A subsequence of the sequence L_jf consists of iterated operators L_qf, i. e. an infinite number of the operators L_jf are identical with L_qf;*

4) *If a non-negative function $T(P)$ which is continuous on the closed region $\bar{R}$ and which vanishes at the boundary points of R attains its maximum value $M (M > 0)$ at a point P_0 and if there is a plane α going through P_0 such that for every point Q of R lying on one side of α we have always $T(Q) < M$, then there exists at least one iterated operator L_qf such that at P_0 we have*

$$L_qT(P_0) < M. \tag{29.1}$$

Under these conditions taking an arbitrary function $f_1(P)$ continuous on the closed region $\bar{R}$ and putting

$$f_2(P) = L_1f_1(P), \cdots, f_{k+1}(P) = L_kf_k(P), \cdots \tag{29.2}$$

the sequence of functions $f_k(P)$ converges uniformly in the closed region $\bar{R}$ towards a function $U(P)$ harmonic in R and having the same boundary values as the original function $f_1(P)$.

30. Nearly all the solutions for Dirichlet's problem which have been proposed since the time of *Schwarz* and *Poincaré* and which are based on a convergent sequence of continuous operators can be treated as special cases of the foregoing theorem.

For the rest of them, as for the solution proposed by *O. D. Kellogg*, iterated operators do not appear at first sight and the conditions of the theorem of the last section must be slightly revised if it is wanted to include also these solutions in our theory. In fact suppose that among the operators $L_j f$ there be no iterated operators but that to each operator $L_j f$ we can add a finite number of operators

$$L_{j1}f, L_{j2}f, \cdots, L_{jp_j}f \tag{30.1}$$

such that we have always

$$L_{jk}(L_j f) \equiv L_j f \qquad (k = 1, 2, \cdots, p_j) \tag{30.2}$$

and that for the sequence of operators which we obtain in including all the operators (30.1) after every $L_j f$ all the conditions of § 29 are fulfilled.

If we remark now that the sequence of functions (29.2) has always the same limit for the original and for the completed system of operators because of the conditions (30.2) we see that the convergence theorem holds for the $L_j f$ themselves.

31. The study of the convergence of our sequence of functions $f_k(P)$ is much more complicated if we do not choose to assume that all the points of the boundary of the region R be regular. In this connection the posthumous memoir of *O. D. Kellogg* about the converse of Gauss' Theorem will be of great help. It contains already the proof of such a convergence in a particular case. One could also think of generalizing the condition 2) of our theorem by replacing the identity by a more general relation which tends towards the former for large values of the index number j. In doing this it would perhaps be possible to include the very important and very interesting constructions of *N. Wiener* (which are discontinuous) in the general frame of our theory. But this exceeds the scope of the present note, which was to show that a fairly general result may be obtained without using any higher concept than that of normal families and that of the continuous convergence of continuous functions. It would though certainly be interesting to investigate more thoroughly, as I have done, the limits at which the theorem of convergence ceases to be true.

PASADENA, CALIFORNIA.

REFERENCES.

H. A. Schwarz, "Zur Integration der partiellen Differentialgleichung $\partial^2 u/\partial x^2 + \partial^2 u/\partial y^2 = 0$," *Journal für reine und angewandte Mathematik*, Bd. 74, pp. 218-253. *Ges. Abh.*, Bd. 2 (1872), pp. 175-210.

H. Poincaré, "Sur les équations aux dérivées partielles de la physique mathématique," *American Journal of Mathematics*, vol. 13 (1890).

H. Lebesgue, "Sur le problème de Dirichlet," *Comptes Rendus* (Paris), vol. 154 (1912), p. 335.

O. Perron, "Eine neue Behandlung der ersten Randwertaufgabe für $\Delta u = 0$," *Mathematische Zeitschrift*, Bd. 18 (1923), pp. 42-54.

T. Radó und F. Riesz, "Über die erste Randwertaufgabe für $\Delta u = 0$," *ibid.*, Bd. 22 (1925), pp. 41-44.

F. Riesz, "Sur les fonctions subharmoniques et leur rapport à la théorie du potentiel (Première Partie)," *Acta Mathematica*, vol. 48 (1926), pp. 329-343.

N. Wiener, "The Dirichlet problem," *Journal of Mathematics and Physics*, Massachusetts Institute of Technology, vol. 3 (1924), pp. 127-146.

O. D. Kellogg, *Foundations of Potential Theory*, Berlin, Springer, 1929.

———, "Converses of Gauss' theorem on the arithmetic mean," *Transactions of the American Mathematical Society*, vol. 36 (1934), pp. 227-242.

C. Carathéodory, "Stetige Konvergenz und normale Familien von Funktionen," *Mathematische Annalen*, Bd. 101 (1929), pp. 515-533.

BOUNDED ANALYTIC FUNCTIONS IN SEVERAL VARIABLES AND MULTIPLE LAPLACE INTEGRALS.*

By S. Bochner.

Introduction. We shall consider the k-dimensional Euclidean space of the real points $x = (x_1, \cdots, x_k)$ and its complex extension consisting of the complex points $z = (z_1, \cdots, z_k)$; $z_\kappa = x_\kappa + iy_\kappa$. Any point set S of the real space is the basis of a *tube* $T = T_S$ of the complex space. The latter is the point set of the complex space consisting of all k-dimensional planes

$$x_\kappa = x_\kappa{}^0; \qquad -\infty < y_\kappa < \infty, \qquad (\kappa = 1, \cdots, k),$$

where $(x_1{}^0, \cdots, x_k{}^0)$ is an arbitrary point of S. A tube is an (open) domain if and only if its basis is a domain. We shall say that a tube T' lies within a tube T if the closure of T' is part of the interior of T.

The transformation

$$w_\kappa = e^{-z_\kappa} \qquad (\kappa = 1, \cdots, k) \tag{1}$$

transforms a tube of the z-space into a so-called Reinhardt set of the w-space and, conversely, the image of a Reinhardt set in the w-space is a tube of the z-space.[1] For shortness, we shall term a Reinhardt region *convex* if the corresponding tube is convex in the usual sense. Any function analytic in a Reinhardt region R has a Laurent expansion

$$\sum \cdot \overset{\infty}{\underset{-\infty}{\cdot}} \cdot \sum a_{n_1 \ldots n_k} w_1{}^{n_1} \cdots w_k{}^{n_k} \tag{2}$$

which is absolutely convergent in R. On the other hand the largest domain of absolute convergence of a series of the form (2) is a convex Reinhardt region.[2] Therefore, if a function is analytic in a Reinhardt region R it also exists in the smallest convex Reinhardt region enclosing R. In z-coördinates we obtain the following result. *If a function $f(z) = f(z_1, \cdots, z_k)$ is analytic in a tube T and has the period $2\pi i$ in each variable then it also exists in the convex closure $\bar{T}$ of T and has the period $2\pi i$ in each variable.*

Similar theorems hold for many classes of analytic functions other than

* Received June 14, 1937.

[1] H. Behnke und P. Thullen, "Theorie der Funktionen mehrerer komplexer Veränderlichen," *Ergebnisse der Mathematik*, vol. 3 (1934), pp. 33-34.

[2] Behnke-Thullen, *ibid.*, pp. 37-38.

periodic. As an illustration we shall discuss the class of bounded functions by two different methods. In the first proof we shall represent functions of integrable square by multiple Laplace integrals of the form

$$\int_{-\infty}^{\infty} \cdots \int \exp(-z_1 t_1 - \cdots - z_k t_k) \cdot \phi(t)\, dv_t, \tag{3}$$

the symbol dv_t denoting the Euclidean volume element $dt_1 \cdots dt_k$.

Functions of integrable square. A function $f(z)$ which is analytic in T_S is of integrable square in T_S if the function

$$f_x(y) \equiv f(x_1 + iy_1, \cdots, x_k + iy_k) \tag{4}$$

belongs to the Lebesgue class L_2 over the y-space, for every $x \subset S$, and moreover,

$$\int_{-\infty}^{\infty} \cdots \int |f_x(y)|^2\, dv_y \leq K \equiv K_S. \tag{5}$$

Of course, if (5) holds only for every domain S' within S then $f(z)$ is of integrable square within T.[3]

Let $\phi(t) = \phi(t_1, \cdots, t_k)$ be a measurable function of the real t-space. If, for every $x \subset S$, the function

$$\phi_x(t) = \exp(-x_1 t_1 - \cdots - x_k t_k) \cdot \phi(t) \tag{6}$$

belongs to L_2 in the t-space then the integral (3) is absolutely and uniformly convergent within T_S and the sum function $f(z)$ is of integrable square within T_S. In proving this we may assume, applying the Heine-Borel theorem, that S is an interval

$$a_\kappa \leq x_\kappa \leq b_\kappa \qquad (\kappa = 1, \cdots, k). \tag{7}$$

Furthermore we may split the range of integration in (3) into the 2^k octants $t_1 \gtrless 0, \cdots, t_k \gtrless 0$. For instance, if we restrict ourselves to the octant $t_1 > 0, \cdots, t_k > 0$, then (3) has the value

$$\int_{-\infty}^{\infty} \cdots \int \exp[-(y_1 t_1 + \cdots + y_k t_k) i] \cdot \phi_x(t)\, dv_t. \tag{8}$$

[3] A function has a property within a tube T if it has the property in every tube T' within T.

The absolute and uniform convergence of (8) follows immediately from the relation

$$\left[\int_0^\infty\cdots\int |\phi_{x+\xi}(t)|\,dv_t\right]^2$$
$$\leq \int_0^\infty\cdots\int \exp(-2\xi_1 t_1 - \cdots - 2\xi_k t_k)\,dv_t\cdot\int_0^\infty\cdots\int |\phi_x(t)|^2\,dv_t$$

which is a consequence of Schwarz' inequality. The sum function of (8) is therefore analytic in $z_1,\cdots,z_k$, and relation (5) follows immediately from the Parseval equation

$$\int_{-\infty}^\infty\cdots\int |f_x(y)|^2\,dv_y = (2\pi)^{k/2}\int_0^\infty\cdots\int |\phi_x(t)|^2\,dv_t.$$

Any other octant can be treated in a similar fashion or can be reduced to the first octant by changing the signs of the negative t-components.

Conversely, any function which is of integrable square in S has a unique representation of the form (3). In fact, by Plancherel's theorem, the function $f_x(y)$ has a representation

$$(9)\qquad f_x(y) \sim \int_{-\infty}^\infty\cdots\int \exp[-(y_1t_1+\cdots+y_kt_k)i]\cdot\phi_x(t)\,dv_t,$$

and we have to prove that (6) holds. Integrating (5) with respect to x over the domain S we obtain

$$\int_T |f_x(y)|^2\,dv_y dv_x < \infty.$$

Hence there exists a constant A such that for any polycylinder

$$P_\zeta:\ |z_\kappa - \zeta_\kappa| \leq r_\kappa$$

which lies entirely in T_S the relation

$$\int_{P_\zeta} |f(z)|^2\,dv_x dv_y \leq A$$

holds. Expanding $f(z)$ in a power series around ζ we easily obtain

$$|f(\zeta)|^2 \leq \frac{1}{\pi^k r_1^2\cdots r_k^2}\int_{P_\zeta} |f(z)|^2\,dv_x dv_y.$$

Thus $f(z)$ is bounded within S. With the Fejer kernel

$$F_n(\eta) = \prod_{\kappa=1}^{k} \frac{(\sin (n/2)\eta_\kappa)^2}{\pi n \eta_\kappa^2}$$

we form the sequence of functions

$$(10) \quad f^n(z) = \int_{-\infty}^{\infty} \cdots \int f(x_1 + iy_1 + i\eta_1 + , \cdots , x_k + iy_k + i\eta_k) F_n(\eta_1, \cdots , \eta_k) dv_\eta.$$

They have the following properties: α) they are analytic and of integrable square within T_S, β) for each (x, y)

$$\lim_{n \to \infty} f_x^n(y) = f_x(y),$$

γ) the Fourier transform $\phi_x^n(t)$ of $f_x^n(t)$, see (9), has the value

$$\phi_x(t) \prod_{\kappa=1}^{k} \left(1 - \frac{|t_\kappa|}{n}\right),$$

if $|t_x| \leq 1$ and the value 0 if some t_κ is > 1.[4] As a consequence of these properties it is sufficient to prove relation (6) for functions $\phi_x(t)$ vanishing outside a finite interval of the t-space. We pick any special point $x^0 \subset S$ and introduce the functions

$$(11) \qquad \phi(t) = \exp(x_1^0 t_1 + \cdots + x_k^0 t_k) \cdot \phi_{x^0}(t)$$

$$g(z) = \int_{-\infty}^{\infty} \cdots \int \exp(-z_1 t_1 - \cdots - z_k t_k) \cdot \phi(t) dv_t.$$

Since $\phi(t)$ vanishes outside a finite interval the function $g(z)$ is analytic. By (11), it is equal to the given function $f(z)$ on the linear manifold $x_1 = x_1^0, \cdots , x_k = x_k^0$. Therefore $f(z) = g(z)$ identically, and this completes the proof of (6).

Finally, the point set S for which $\phi_x(t)$ belongs to L_2 in the t-space is convex. In fact, if x^1 and x^2 are two points of S, and $x = \alpha x^1 + (1 - \alpha) x^2$, $0 < \alpha < 1$, is a point of the segment joining these points, then the quantities

$$E_1 = \exp(-x_1^1 t_1 - \cdots - x_k^1 t_k), \qquad E_2 = \exp(-x_1^2 t_1 - \cdots - x_k^2 t_k) \cdot$$
$$E = \exp(-x_1 t_1 - \cdots - x_k t_k)$$

[4] S. Bochner, *Vorlesungen über Fouriersche Integrale*, 1932, p. 191, Satz 57.

satisfy the relation

$$E = E_1{}^a E_2{}^{1-a} \leq \max(E_1, E_2) \leq E_1 + E_2.$$

Hence we conclude that *a function which is analytic and of integrable square within a tube exists and is of integrable square within the convex closure of the tube.*

Bounded functions. From the last statement we shall conclude that *a function which is analytic and bounded within a tube exists and is bounded within its convex closure.*

For the proof we may assume that S is a bounded domain whose closure is contained in the octant $x_1 > 0, \cdots, x_k > 0$. This being the case, the function

$$g(z) = \frac{f(z)}{z_1 \cdots z_k} \tag{12}$$

is of integrable square within T and therefore it exists and is bounded within $\bar{T}$. Thus $f(z)$ is bounded within T and $O(z_1 \cdots z_k)$ within T. Given any two points x^1 and x^2 of S and any real numbers $\eta_1, \cdots, \eta_k$ we form the function of the one complex variable $w = u + iv$,

$$h(w) = f(x_1{}^1 + w(x_1{}^2 - x_1{}^1) + i\eta_1, \cdots, x_k + w(x_k{}^2 - x_k{}^1) + i\eta_k).$$

It is defined in the strip $0 \leq u \leq 1$, it is bounded on the boundary $u = 0$, $u = 1$, and it is $O(v^k)$ as $v \to \pm \infty$. By Phragmen-Lindelöf, $h(w)$ is bounded in the strip, and any upper bound for its values on the boundary is a bound for its values in the interior of the strip. For appropriate values of x^1, x^2, η_k, and w every point of $\bar{T}$ can be reached, and hence we easily conclude that $f(z)$ is bounded within $\bar{T}$.

Functions in octant-shaped tubes. If a Reinhardt region contains the origin and is star-shaped with respect to the origin then the corresponding tube is "octant-shaped": if $(x_1{}^0, \cdots, x_k{}^0)$ is a point of its base S then the whole octant

$$x_\kappa \leq x_\kappa{}^0 \qquad (\kappa = 1, \cdots, k) \tag{13}$$

is contained in S. The convex closure of an octant-shaped tube is again octant-shaped.

If $f(z)$ is analytic in an octant-shaped tube T it also exists in the convex closure $\bar{T}$ of T; if it is bounded within T it is also bounded within $\bar{T}$.

Let $x' = (x'_1, \cdots, x'_k)$, $x'' = (x'_1, \cdots, x''_k)$ be any two points of S. We want to show that $f(z)$ is analytic for the points

$$(14) \qquad x_\kappa \leq \alpha x'_\kappa + (1-\alpha) x''_\kappa, \qquad 0 \leq \alpha \leq 1$$

of the basis S of T.

For $\epsilon > 0$ sufficiently small and $A > 0$ sufficiently large the two polycylinders

$$(15) \qquad \begin{aligned} P'&: \quad |z_\kappa + A| \leq x'_\kappa + 2\epsilon + A \\ P''&: \quad |z_\kappa + A| \leq x''_\kappa + 2\epsilon + A \end{aligned}$$

are both contained in T. The smallest convex Reinhardt region (*not* tube) of the z-space which is concentric with, and contains P' and P'', contains the polycylinders

$$(16) \qquad |z_\kappa + A| \leq (x'_\kappa + \epsilon + A)^\alpha (x''_\kappa + \epsilon + A)^{1-\alpha}.$$

The sum of these point-sets for variable values of A contains the points of S for which

$$x_\kappa < \alpha(x'_\kappa + \epsilon) + (1-\alpha)(x''_\kappa + \epsilon)$$

and therefore $f(z)$ is analytic in (14). The complex variables $z_1, \cdots, z_k$ can be altered by arbitrary fixed quantities $iy_1{}^0, \cdots, iy_k{}^0$; therefore $f(z)$ is analytic for all points whose real parts satisfy relation (14). The points x', x'' being arbitrary points of S we thus conclude that $f(z)$ is analytic throughout $\tilde{T}$. This proves the first part of the theorem. The second part is a consequence of the following remark.[5] If $|f(z)|$ is $\leq M$ in (15) then it is $\leq M \cdot C$ in (16) where C depends only on x'_κ, x''_κ, ϵ, A.

Application to bounded functions. We shall give a new proof of our previous statement concerning bounded functions, and we shall again assume that our function is a function of integrable square, see (12).

If a function $f(z)$ of one variable is of integrable square in the closed interval

$$a \leq x \leq b$$

then Cauchy's formula

$$f(z) = \frac{1}{2\pi i} \int_{a+i\infty}^{a-i\infty} \frac{f(s)}{s-z}\, dx + \frac{1}{2\pi i} \int_{b-i\infty}^{b+i\infty} \frac{f(s)}{s-z}\, ds$$

[5] The remark can be proved easily by the reasoning in H. Tietze, "Über den Bereich absoluter Konvergenz von Potenzreihen mehrerer Veränderlichen," *Mathematische Annalen* 99 (1928), p. 181.

leads to a decomposition of $f(z)$ into a sum of two functions $f_1(z)$, $f_2(z)$ of which $f_1(z)$ is analytic and bounded within $x > a$ and $f_2(z)$ is analytic and bounded within $x < b$. Up to additive constants such a decomposition is unique.[6] More generally, if $f(z) = f(z_1, \cdots, z_k)$ is of integrable square in the tube (7), it can be written as a sum of 2^k functions $f_\nu(z)$, $(\nu = 1, \cdots, 2^k)$, each of which is analytic and bounded within an octant containing the interior of the tube (7). For example, if $k = 2$, the four octants are 1) $x_1 > a_1$, $x_2 > a_2$, 2) $x_1 > a_1, x_2 < b_2$, 3) $x_1 < b_1, x_2 > a_2$, 4) $x_1 < b_1, x_2 < b$. The functions $f_\nu(z)$ are again uniquely determined up to additive constants.

We cover the basis S by a denumerable number of octants

$$a_\kappa{}^n \leq x_\kappa \leq b_\kappa{}^n, \qquad (n = 1, 2, \cdots).$$

By the uniqueness property the functions $f_\nu{}^n(z)$ can be so normalized that for fixed ν they are analytic continuations of each other. Therefore $f(z)$ is a sum of functions $f_\nu(z)$ each of which is analytic and bounded within an octant-shaped tube T_ν containing T. By the result of the previous section each $f_\nu(z)$ is analytic, and bounded within the convex closure $\bar{T}_\nu$ of T_ν. But the intersection of the domains $\bar{T}_\nu$ contains the convex closure $\bar{T}$ of T.

PRINCETON UNIVERSITY.

[6] Harald Bohr, "Fastperiodische Funktionen III," *Acta Mathem.*, Bd. 47 (1926), pp. 250-251.

THE CHARACTERS OF THE SYMMETRIC GROUP.*

By F. D. Murnaghan.

We have given in a recent paper (**1**) a recurrence formula (generalising the classical formula of Schur) which makes it possible to determine the characters of the symmetric group on n letters when those of the symmetric groups on a lesser number m letters $(m = 1, 2, \cdots, n-1)$ are known. This method, whilst quite convenient for a systematic construction of the character tables of the symmetric groups, labors under the handicap of all recurrence methods: if one wishes to know the characters of a particular symmetric group (say $n = 20$) one must first know the characters of the symmetric groups on $m = 1, 2, \cdots 19$ letters. Explicit and convenient formulae for the characters of *certain* classes (e. g. the identity class and the transposition class) are available but the number of these classes is very restricted; thus for the class $\alpha_1 = n - 4$, $\alpha_2 = 2$ (i. e. the class consisting of two transpositions, which is important in the applications to nuclear physics) it is not easy to give a convenient explicit formula. The best one can do at present is to use the recurrence formula mentioned above together with the explicit formula of Frobenius for the characters of the transposition class $(\alpha_1 = m - 2, \alpha'_2 = 1)$ of the symmetric group on $m = n - 2$ letters. It is, however, possible to attack the problem of furnishing explicit formulae for the characters of the symmetric group from another side: instead of furnishing formulae for the characters of a given *class* (the *representation* varying) we may give formulae for the characters of a given *representation* (the *class* varying). This point of view has certain advantages when we are concerned with the applications to nuclear physics; for, there, only certain representations (namely, those corresponding to partitions of n into not more than four parts) play a rôle. Certain formulae of this type were given by Frobenius (**2**) and these were extended by Littlewood and Richardson (**3**). The method followed by the last named writers is somewhat complicated and it is the purpose of the present paper to show that the formulae they give may be readily obtained by the methods of our previous paper and to further extend their results. We shall suppose the reader to be acquainted with the notation and methods of the paper (**1**) and shall refer to this paper, where necessary, by merely giving the page number.

* Received August 30, 1937.

1. The characters of the two-element-partition representations $D(\lambda_1, \lambda_2)$. If $(\lambda) = (\lambda_1, \lambda_2, \cdots, \lambda_k)$, $\lambda_1 \geqq \lambda_2 \geqq \cdots \geqq \lambda_k > 0$ is any partition of n the characteristic of the reducible representation $\Delta(\lambda)$ of the symmetric group on n letters is the product $q_{\lambda_1}(s) \cdots q_{\lambda_k}(s)$ where the $q_{\lambda_j}(s)$, $j = 1, 2, \cdots, k$, are the various principal characteristics (pp. 444-445). From the formula for $q_n(s)$: $q_n(s) = \sum_{(\alpha)} \frac{1}{\alpha_1 ! \, \alpha_2 ! \cdots \alpha_n !} \left(\frac{s_1}{1}\right)^{\alpha_1} \cdots \left(\frac{s_n}{n}\right)^{\alpha_n}$ it is clear that the coefficient of $\frac{1}{\alpha_1 ! \cdots \alpha_n !} \left(\frac{s_1}{1}\right)^{\alpha_1} \cdots \left(\frac{s_n}{n}\right)^{\alpha_n}$ in the product $s_1 \frac{\partial q_n}{\partial s_1}$ (or, equivalently, $s_1 q_{n-1}$, p. 445) is α_1. Equally evident is the fact that the coefficient of the same expression in the product $s_1{}^2 \frac{\partial^2 q_n}{\partial s_1{}^2}$ (or, equivalently, $s_1{}^2 q_{n-2}$) is $\alpha_1(\alpha_1 - 1)$; and, generally, the coefficient of the expression referred to in $s_1{}^{\beta_1} q_{n-\beta_1}$ is $\alpha_1(\alpha_1 - 1) \cdots (\alpha_1 - \beta_1 + 1)$. Similarly the coefficient of the same expression in $\left(\frac{s_2}{2}\right)^{\beta_2} q_{n-2\beta_2}$ is $\alpha_2(\alpha_2 - 1) \cdots (\alpha_2 - \beta_2 + 1)$; and, generally, the coefficient of $\frac{1}{\alpha_1 ! \cdots \alpha_n !} \left(\frac{s_1}{1}\right)^{\alpha_1} \cdots \left(\frac{s_n}{n}\right)^{\alpha_n}$ in $\left(\frac{s_j}{j}\right)^{\beta_j} q_{n-j\beta_j}$ is $\alpha_j(\alpha_j - 1) \cdots (\alpha_j - \beta_j + 1)$. Hence the coefficient of this same expression in $\left(\frac{s_1}{1}\right)^{\beta_1} \cdots \left(\frac{s_n}{n}\right)^{\beta_n} q_{n-\lambda_2}$, where $\lambda_2 = \beta_1 + 2\beta_2 + \cdots + n\beta_n$, is the product

$$\prod_{j=1}^{n} \{\alpha_j(\alpha_j - 1) \cdots (\alpha_j - \beta_j + 1)\}.$$

Since

$$q_{\lambda_2}(s) = \sum_{(\beta)} \frac{1}{\beta_1 ! \cdots \beta_n !} \left(\frac{s_1}{1}\right)^{\beta_1} \cdots \left(\frac{s_n}{n}\right)^{\beta_n}$$

it follows that the coefficient of

$$\frac{1}{\alpha_1 ! \, \alpha_2 ! \cdots \alpha_n !} \left(\frac{s_1}{1}\right)^{\alpha_1} \cdots \left(\frac{s_n}{n}\right)^{\alpha_n}$$

in the product $q_{\lambda_1}(s) q_{\lambda_2}(s)$, where $\lambda_1 + \lambda_2 = n$, is the summation over all classes (β) of the symmetric group on λ_2 letters of the product $\binom{\alpha_1}{\beta_1} \cdots \binom{\alpha_n}{\beta_n}$ $\left(\text{where } \binom{a}{b} \text{ is the binomial coefficient } \frac{a!}{\overline{a-b}! \, b!}\right)$. In other words the character of the class (α) in the reducible representation $\Delta(\lambda_1, \lambda_2)$ is the sum $\sum_{(\beta)} \binom{\alpha_1}{\beta_1} \cdots \binom{\alpha_n}{\beta_n}$. Since the characteristic of the irreducible representation $D(\lambda_1, \lambda_2)$ is the two rowed determinant $\begin{vmatrix} q_{\lambda_1}(s) \, q_{\lambda_1+1}(s) \\ q_{\lambda_2-1}(s) \, q_{\lambda_2}(s) \end{vmatrix}$ (p. 460) we have

$D(\lambda_1, \lambda_2) = \Delta(\lambda_1, \lambda_2) - \Delta(\lambda_1 + 1, \lambda_2 - 1)$ and so the character of the class (α) in the irreducible representation $D(\lambda_1, \lambda_2)$ is

$$\sum_{(\beta)} \binom{\alpha_1}{\beta_1} \cdots \binom{\alpha_n}{\beta_n} - \sum_{(\gamma)} \binom{\alpha_1}{\gamma_1} \cdots \binom{\alpha_n}{\gamma_n}$$

the summations being over all classes (β) and (γ) of the symmetric groups on λ_2 and $\lambda_2 - 1$ letters, respectively. Corresponding to each class (γ) of the symmetric group on $\lambda_2 - 1$ letters is the class $\beta_1 = \gamma_1 + 1$, $\beta_2 = \gamma_2, \cdots,$ $\beta_n = \gamma_n$ of the symmetric group on λ_2 letters and so we may write our formula for the character of the class (α) in the irreducible representation $D(\lambda_1, \lambda_2)$ as follows:

$$(1) \qquad D(\lambda_1, \lambda_2)_{(\alpha)} = \sum_{(\gamma)} \frac{1}{\gamma_1 + 1!} \alpha_1(\alpha_1 - 1) \cdots$$
$$\cdots (\alpha_1 - \gamma_1 + 1)(\alpha_1 - 2\gamma_1 - 1) \binom{\alpha_2}{\gamma_2} \cdots \binom{\alpha_n}{\gamma_n} + \sum_{(\beta)} \binom{\alpha_2}{\beta_2} \cdots \binom{\alpha_n}{\beta_n}$$

the summation in the first term being over all classes (γ) of the symmetric group on $\lambda_2 - 1$ letters and in the second over all *those classes* (β) *of the symmetric group on* λ_2 *letters which do not contain any unary cycle* (= *fixed letter*).

As examples of the use of this formula we list the following:

$D(n-1, 1)_{(\alpha)} = \alpha_1 - 1$ (this degenerate case being best read off directly from the relation $\{n-1, 1\} = q_{n-1}q_1 - q_n$)

$$D(n-2, 2)_{(\alpha)} = \tfrac{1}{2}\alpha_1(\alpha_1 - 3) + \alpha_2$$
$$D(n-3, 3)_{(\alpha)} = \tfrac{1}{6}\alpha_1(\alpha_1 - 1)(\alpha_1 - 5) + (\alpha_1 - 1)\alpha_2 + \alpha_3$$
$$D(n-4, 4)_{(\alpha)} = \tfrac{1}{24}\alpha_1(\alpha_1 - 1)(\alpha_1 - 2)(\alpha_1 - 7) + \tfrac{1}{2}\alpha_1(\alpha_1 - 3)\alpha_2 + (\alpha_1 - 1)\alpha_3 + \tfrac{1}{2}\alpha_2(\alpha_2 - 1) + \alpha_4$$
$$D(n-5, 5)_{(\alpha)} = \tfrac{1}{120}\alpha_1(\alpha_1 - 1)(\alpha_1 - 2)(\alpha_1 - 3)(\alpha_1 - 9) + \tfrac{1}{6}\alpha_1(\alpha_1 - 1)(\alpha_1 - 5)\alpha_2 + \tfrac{1}{2}\alpha_1(\alpha_1 - 3)\alpha_3 + \tfrac{1}{2}(\alpha_1 - 1)\alpha_2(\alpha_2 - 1) + (\alpha_1 - 1)\alpha_4 + \alpha_2\alpha_3 + \alpha_5$$
$$D(n-6, 6)_{(\alpha)} = \tfrac{1}{720}\alpha_1(\alpha_1 - 1)(\alpha_1 - 2)(\alpha_1 - 3)(\alpha_1 - 4)(\alpha_1 - 11) + \tfrac{1}{24}\alpha_1(\alpha_1 - 1)(\alpha_1 - 2)(\alpha_1 - 7)\alpha_2 + \tfrac{1}{6}\alpha_1(\alpha_1 - 1)(\alpha_1 - 5)\alpha_3 + \tfrac{1}{4}\alpha_1(\alpha_1 - 3)\alpha_2(\alpha_2 - 1) + \tfrac{1}{2}\alpha_1(\alpha_1 - 3)\alpha_4 + (\alpha_1 - 1)\alpha_2\alpha_3 + (\alpha_1 - 1)\alpha_5 + \tfrac{1}{6}\alpha_2(\alpha_2 - 1)(\alpha_2 - 2) + \alpha_2\alpha_4 + \tfrac{1}{2}\alpha_3(\alpha_3 - 1) + \alpha_6$$
$$D(n-7, 7)_{(\alpha)} = (1/7!)\alpha_1(\alpha_1 - 1)(\alpha_1 - 2)(\alpha_1 - 3)(\alpha_1 - 4)(\alpha_1 - 5)(\alpha_1 - 13) + \tfrac{1}{120}\alpha_1(\alpha_1 - 1)(\alpha_1 - 2)(\alpha_1 - 3)(\alpha_1 - 9)\alpha_2 + \tfrac{1}{24}\alpha_1(\alpha_1 - 1)(\alpha_1 - 2)(\alpha_1 - 7)\alpha_3 + \tfrac{1}{6}\alpha_1(\alpha_1 - 1)(\alpha_1 - 5)\alpha_4 +$$ (see next page)

$$
\begin{aligned}
&+ \tfrac{1}{12}\alpha_1(\alpha_1-1)(\alpha_1-5)\alpha_2(\alpha_2-1) + \tfrac{1}{2}\alpha_1(\alpha_1-3)\alpha_2\alpha_3 \\
&+ \tfrac{1}{2}\alpha_1(\alpha_1-3)\alpha_5 + (\alpha_1-1)\alpha_6 + (\alpha_1-1)\alpha_2\alpha_4 \\
&+ \tfrac{1}{6}(\alpha_1-1)\alpha_2(\alpha_2-1)(\alpha_2-2) \\
&+ \tfrac{1}{2}(\alpha_1-1)\alpha_3(\alpha_3-1) + \alpha_2\alpha_5 + \tfrac{1}{2}\alpha_2(\alpha_2-1)\alpha_3 \\
&+ \alpha_3\alpha_4 + \alpha_7.
\end{aligned}
$$

$$
\begin{aligned}
D(n-8,8)_{(\alpha)} = {} & (1/8\,!)\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-4)(\alpha_1-5)(\alpha_1-6)(\alpha_1-15) \\
&+ (1/6\,!)\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-4)(\alpha_1-11)\alpha_2 \\
&+ (1/5\,!)\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-9)\alpha_3 \\
&+ \tfrac{1}{24}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-7)\alpha_4 \\
&+ \tfrac{1}{48}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-7)\alpha_2(\alpha_2-1) \\
&+ \tfrac{1}{6}\alpha_1(\alpha_1-1)(\alpha_1-5)\alpha_2\alpha_3 + \tfrac{1}{6}\alpha_1(\alpha_1-1)(\alpha_1-5)\alpha_5 \\
&+ \tfrac{1}{2}\alpha_1(\alpha_1-3)\alpha_6 + \tfrac{1}{2}\alpha_1(\alpha_1-3)\alpha_2\alpha_4 \\
&+ \tfrac{1}{12}\alpha_1(\alpha_1-3)\alpha_2(\alpha_2-1)(\alpha_2-2) + \tfrac{1}{4}\alpha_1(\alpha_1-3)\alpha_3(\alpha_3-1) \\
&+ (\alpha_1-1)\alpha_2\alpha_5 + \tfrac{1}{2}(\alpha_1-1)\alpha_2(\alpha_2-1)\alpha_3 + (\alpha_1-1)\alpha_3\alpha_4 \\
&+ (\alpha_1-1)\alpha_7 + \tfrac{1}{24}\alpha_2(\alpha_2-1)(\alpha_2-2)(\alpha_2-3) \\
&+ \tfrac{1}{2}\alpha_2\alpha_3(\alpha_3-1) + \tfrac{1}{2}\alpha_2(\alpha_2-1)\alpha_4 + \alpha_2\alpha_6 + \alpha_3\alpha_5 \\
&+ \tfrac{1}{2}\alpha_4(\alpha_4-1) + \alpha_8.
\end{aligned}
$$

It is worthy of note that these *formulae* are the same for all values of n; their *evaluations*, of course, depend on n since (α) is a class of the symmetric group on n letters, i. e. $\alpha_1 + 2\alpha_2 + \cdots + n\alpha_n = n$. The formula for $D(\lambda_1, \lambda_2)_{(\alpha)}$ involves only the $\alpha_1, \cdots, \alpha_{\lambda_2}$ being independent of the α_j for which $j > \lambda_2$. The first term in each formula yields, when α_1 is put equal to n, Frobenius' formula for the dimension of the representation $D(\lambda_1, \lambda_2)$. As a check on the accuracy of the formulae (e. g., to detect a possible error in copying) we may observe that $D(\lambda_1, \lambda_2) = 0$ when $\lambda_1 = \lambda_2 - 1$, i. e. when $n = 2\lambda_2 - 1$; hence the expression for $D(\lambda_1, \lambda_2)_{(\alpha)}$ must vanish when α is *any* class of $2\lambda_2 - 1$. E. g., the polynomial in $(\alpha_1, \alpha_2, \alpha_3)$ giving $D(n-3, 3)_{(\alpha)}$ must vanish when $(\alpha_1, \alpha_2, \alpha_3)$ have the following sets of values: $(5,0,0)$, $(3,1,0)$, $(2,0,1)$, $(1,2,0)$, $(0,1,1)$. Since the characters of any representation follow from those of the associated representation by a mere change of sign of the characters of the odd classes the formulae derived above furnish the characters of all representations which correspond to partitions of n in which no element > 2.

2. The characters of the three-element-partition representations $D(\lambda_1, \lambda_2, \lambda_3)$. On expanding the characteristic

$$
\{\lambda_1, \lambda_2, \lambda_3\} = \begin{vmatrix} q_{\lambda_1}(s) & q_{\lambda_1+1}(s) & q_{\lambda_1+2}(s) \\ q_{\lambda_2-1}(s) & q_{\lambda_2}(s) & q_{\lambda_2+1}(s) \\ q_{\lambda_3-2}(s) & q_{\lambda_3-1}(s) & q_{\lambda_3}(s) \end{vmatrix}
$$

of $D(\lambda_1, \lambda_2, \lambda_3)$ in terms of the first row and observing that

$$\begin{vmatrix} q_{\lambda_2-1}(s) & q_{\lambda_2+1}(s) \\ q_{\lambda_3-2}(s) & q_{\lambda_3}(s) \end{vmatrix} = \begin{vmatrix} q_{\lambda_2-1}(s) & q_{\lambda_2}(s) \\ q_{\lambda_3-1}(s) & q_{\lambda_3}(s) \end{vmatrix} + \begin{vmatrix} q_{\lambda_2}(s) & q_{\lambda_2+1}(s) \\ q_{\lambda_3-2}(s) & q_{\lambda_3-1}(s) \end{vmatrix}$$

$$= \{\lambda_2 - 1, \lambda_3\} + \{\lambda_2, \lambda_3 - 1\} = \frac{\partial}{\partial s_1}\{\lambda_2, \lambda_3\}$$

we find

$$\{\lambda_1, \lambda_2, \lambda_3\} = q_{\lambda_1}(s)\{\lambda_2, \lambda_3\} - q_{\lambda_1+1}(s)\frac{\partial}{\partial s_1}\{\lambda_2, \lambda_3\} + q_{\lambda_1+2}(s)\{\lambda_2 - 1, \lambda_3 - 1\}.$$

If, now, the characters of the symmetric groups on $\lambda_2 + \lambda_3$ and $\lambda_2 + \lambda_3 - 2$ letters are known we write

$$\{\lambda_2, \lambda_3\} = \sum_{(\beta)} \frac{\{\lambda_2, \lambda_3\}_{(\beta)}}{\beta_1! \cdots \beta_n!} \left(\frac{s_1}{1}\right)^{\beta_1} \cdots \left(\frac{s_n}{n}\right)^{\beta_n}$$

the summation being over all classes (β) of the symmetric group on $\lambda_2 + \lambda_3$ letters. Hence the coefficient of

$$\frac{1}{\alpha_1! \cdots \alpha_n!} \left(\frac{s_1}{1}\right)^{\alpha_1} \cdots \left(\frac{s_n}{n}\right)^{\alpha_n}.$$

in $q_{\lambda_1}(s)\{\lambda_2, \lambda_3\}$ is

$$\sum_{(\beta)} \{\lambda_2, \lambda_3\}_{(\beta)} \binom{\alpha_1}{\beta_1} \cdots \binom{\alpha_n}{\beta_n}$$

Also

$$\frac{\partial}{\partial s_1}\{\lambda_2, \lambda_3\} = \sum_{(\beta')} \frac{\{\lambda_2, \lambda_3\}_{(\beta)}}{\beta'_1!\, \beta_2! \cdots \beta_n!} \left(\frac{s_1}{1}\right)^{\beta'_1} \left(\frac{s_2}{2}\right)^{\beta_2} \cdots \left(\frac{s_n}{n}\right)^{\beta_n}$$

where $\beta'_1 = \beta_1 - 1$ so that $(\beta') = (\beta'_1, \beta_2, \cdots, \beta_n)$ is an arbitrary class of the symmetric group on $\lambda_2 + \lambda_3 - 1$ letters; hence the coefficient of

$$\frac{1}{\alpha_1! \cdots \alpha_n!} \left(\frac{s_1}{1}\right)^{\alpha_1} \cdots \left(\frac{s_n}{n}\right)^{\alpha_n}$$

in the product $q_{\lambda_1+1}(s)\dfrac{\partial}{\partial s_1}\{\lambda_2, \lambda_3\}$ is

$$\sum_{(\beta')} \{\lambda_2, \lambda_3\}_{(\beta)} \binom{\alpha_1}{\beta'_1}\binom{\alpha_2}{\beta_2} \cdots \binom{\alpha_n}{\beta_n}$$

the summation being over all classes (β') of the symmetric group on $\lambda_2 + \lambda_3 - 1$ letters and the coefficients $\{\lambda_2, \lambda_3\}_{(\beta)}$ being the characters of $D(\lambda_2, \lambda_3)$ corresponding to those classes of the symmetric group on $\lambda_2 + \lambda_3$ letters which contain *at least one* fixed letter. Finally the coefficient of

$$\frac{1}{\alpha_1! \cdots \alpha_n!} \left(\frac{s_1}{1}\right)^{\alpha_1} \cdots \left(\frac{s_n}{n}\right)^{\alpha_n}$$

in the product $q_{\lambda_1+2}(s)\{\lambda_2-1, \lambda_3-1\}$ is

$$\sum_{(\gamma)} \{\lambda_2-1, \lambda_3-1\}_{(\gamma)} \binom{\alpha_1}{\gamma_1} \cdots \binom{\alpha_n}{\gamma_n}$$

the summation being over all classes (γ) of the symmetric group on $\lambda_2+\lambda_3-2$ letters. Hence we have the result

$$\{\lambda_1, \lambda_2, \lambda_3\}_{(\alpha)} = \sum_{(\beta)} \{\lambda_2, \lambda_3\}_{(\beta)} \binom{\alpha_1}{\beta_1} \cdots \binom{\alpha_n}{\beta_n} - \sum_{(\beta')} \{\lambda_2, \lambda_3\}_{(\beta)} \binom{\alpha_1}{\beta'_1} \cdots \binom{\alpha_n}{\beta_n} + \sum_{(\gamma)} \{\lambda_2-1, \lambda_3-1\}_{(\gamma)} \binom{\alpha_1}{\gamma_1} \cdots \binom{\alpha_n}{\gamma_n}.$$

The difference of the first two terms may be written as

$$\sum_{(\beta')} \frac{1}{\beta_1!} \{\lambda_2, \lambda_3\}_{(\beta)} \alpha_1(\alpha_1-1) \cdots (\alpha_1-\beta'_1+1)(\alpha_1-2\beta'_1-1) \binom{\alpha_2}{\beta_2} \cdots \binom{\alpha_n}{\beta_n} + \sum_{(\delta)} \{\lambda_2, \lambda_3\}_{(\delta)} \binom{\alpha_2}{\delta_2} \cdots \binom{\alpha_n}{\delta_n}$$

where the summation with respect to (δ) is over all classes (δ) of the symmetric group on $\lambda_2+\lambda_3$ letters *which do not contain any unary cycle* (= fixed letter). Hence we have the three-element-partition formula:

$$(2) \quad D(\lambda_1, \lambda_2, \lambda_3)_{(\alpha)}$$
$$= \sum_{(\beta')} \frac{1}{\beta_1!} \{\lambda_2, \lambda_3\}_{(\beta)} \alpha_1(\alpha_1-1) \cdots (\alpha_1-\beta'_1+1)(\alpha_1-2\beta'_1-1) \binom{\alpha_2}{\beta_2} \cdots \binom{\alpha_n}{\beta_n} + \sum_{(\delta)} \{\lambda_2, \lambda_3\}_{(\delta)} \binom{\alpha_2}{\delta_2} \cdots \binom{\alpha_n}{\delta_n} + \sum_{(\gamma)} \{\lambda_2-1, \lambda_3-1\}_{(\gamma)} \binom{\alpha_1}{\gamma_1} \cdots \binom{\alpha_n}{\gamma_n}.$$

As an illustration of the application of this formula we consider the simplest case: $D(n-2, 1^2)$. Here $\lambda_2=1$, $\lambda_3=1$ so that $\lambda_2+\lambda_3-1=1$ and there is only one class (β'), namely, $\beta'_1=1$. There is only one class (δ), namely, $\delta_2=1$ and the last term yields unity. Hence

$$D(n-2, 1^2)_{(\alpha)} = \tfrac{1}{2}\alpha_1(\alpha_1-3) - \alpha_2 + 1 = \tfrac{1}{2}(\alpha_1-1)(\alpha_1-2) - \alpha_2.$$

As examples of the use of (2) we list the following:

$$\begin{aligned}
D(n-2, 1^2)_{(\alpha)} &= \tfrac{1}{2}(\alpha_1-1)(\alpha_1-2) - \alpha_2 \\
D(n-3, 2, 1)_{(\alpha)} &= \tfrac{1}{3}\alpha_1(\alpha_1-2)(\alpha_1-4) - \alpha_3 \\
D(n-4, 3, 1)_{(\alpha)} &= \tfrac{1}{8}\alpha_1(\alpha_1-1)(\alpha_1-3)(\alpha_1-6) \\
&\quad + \tfrac{1}{2}\alpha_1(\alpha_1-3)\alpha_2 - \tfrac{1}{2}\alpha_2(\alpha_2-3) - \alpha_4 \\
D(n-4, 2^2)_{(\alpha)} &= \tfrac{1}{12}\alpha_1(\alpha_1-1)(\alpha_1-4)(\alpha_1-5) - (\alpha_1-1)\alpha_3 \\
&\quad + \alpha_2(\alpha_2-2)
\end{aligned}$$

$$D(n-5,4,1)_{(a)} = \tfrac{1}{30}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-4)(\alpha_1-8) + \tfrac{1}{3}\alpha_1(\alpha_1-2)(\alpha_1-4)\alpha_2 + \tfrac{1}{2}\alpha_1(\alpha_1-3)\alpha_3-(\alpha_2-1)\alpha_3-\alpha_5$$

$$D(n-5,3,2)_{(a)} = \tfrac{1}{24}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-5)(\alpha_1-7) + \tfrac{1}{6}\alpha_1(\alpha_1-1)(\alpha_1-5)\alpha_2-\tfrac{1}{2}(\alpha_1-1)(\alpha_1-2)\alpha_3 - (\alpha_1-1)\alpha_4+\tfrac{1}{2}(\alpha_1-1)\alpha_2(\alpha_2-1)+\alpha_2\alpha_3$$

$$D(n-6,5,1)_{(a)} = \tfrac{1}{144}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-5)(\alpha_1-10) + \tfrac{1}{8}\alpha_1(\alpha_1-1)(\alpha_1-3)(\alpha_1-6)\alpha_2 + \tfrac{1}{3}\alpha_1(\alpha_1-2)(\alpha_1-4)\alpha_3+\tfrac{1}{2}(\alpha_1-1)(\alpha_1-2)\alpha_4 + \tfrac{1}{4}(\alpha_1-1)(\alpha_1-2)\alpha_2(\alpha_2-1) - \tfrac{1}{6}\alpha_2(\alpha_2-1)(\alpha_2-2)-\alpha_2\alpha_4-\tfrac{1}{2}\alpha_3(\alpha_3-1)-\alpha_6$$

$$D(n-6,4,2)_{(a)} = \tfrac{1}{80}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-6)(\alpha_1-9) + \tfrac{1}{8}\alpha_1(\alpha_1-1)(\alpha_1-3)(\alpha_1-6)\alpha_2 - \tfrac{1}{2}(\alpha_1-1)(\alpha_1-2)\alpha_4+\tfrac{1}{4}\alpha_1(\alpha_1-3)\alpha_2(\alpha_2-1) -(\alpha_1-1)\alpha_5+\tfrac{1}{2}\alpha_2(\alpha_2-1)(\alpha_2-3)+\alpha_2\alpha_4$$

$$D(n-6,3^2)_{(a)} = \tfrac{1}{144}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-7)(\alpha_1-8) + \tfrac{1}{24}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-7)\alpha_2 - \tfrac{1}{6}\alpha_1(\alpha_1-1)(\alpha_1-5)\alpha_3-\tfrac{1}{2}\alpha_1(\alpha_1-3)\alpha_4 + \tfrac{1}{4}\alpha_1(\alpha_1-3)\alpha_2(\alpha_2-1)+\alpha_1(\alpha_2-1)\alpha_3 - \tfrac{1}{2}\alpha_2(\alpha_2-1)(\alpha_2-4)-\alpha_2\alpha_3-\alpha_2\alpha_4+\alpha_3(\alpha_3-1)$$

$$D(n-7,6,1)_{(a)} = \tfrac{1}{840}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-4)(\alpha_1-6)(\alpha_1-12) + \tfrac{1}{30}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-4)(\alpha_1-8)\alpha_2 + \tfrac{1}{8}\alpha_1(\alpha_1-1)(\alpha_1-3)(\alpha_1-6)\alpha_3 + \tfrac{1}{3}\alpha_1(\alpha_1-2)(\alpha_1-4)\alpha_4 + \tfrac{1}{6}\alpha_1(\alpha_1-2)(\alpha_1-4)\alpha_2(\alpha_2-1) + \tfrac{1}{2}(\alpha_1-1)(\alpha_1-2)\alpha_2\alpha_3 + \tfrac{1}{2}(\alpha_1-1)(\alpha_1-2)\alpha_5 - \tfrac{1}{2}\alpha_2(\alpha_2-1)\alpha_3-\alpha_2\alpha_5-\alpha_3\alpha_4-\alpha_7$$

$$D(n-7,5,2)_{(a)} = \tfrac{1}{360}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-4)(\alpha_1-7)(\alpha_1-11) + \tfrac{1}{20}\alpha_1(\alpha_1-1)(\alpha_1-3)(\alpha_1-4)(\alpha_1-7)\alpha_2 + \tfrac{1}{12}\alpha_1(\alpha_1-1)(\alpha_1-4)(\alpha_1-5)\alpha_3 + \tfrac{1}{6}\alpha_1(\alpha_1-1)(\alpha_1-5)\alpha_2(\alpha_2-2)-\tfrac{1}{2}\alpha_1(\alpha_1-3)\alpha_5 + \tfrac{1}{3}(\alpha_1-1)\alpha_2(\alpha_2-1)(\alpha_2-2) - \tfrac{1}{2}(\alpha_1-1)\alpha_3(\alpha_3-1)-(\alpha_1-1)\alpha_6 + \alpha_2(\alpha_2-2)\alpha_3+(\alpha_2-1)\alpha_5$$

$$D(n-7,4,3)_{(a)} = \tfrac{1}{360}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-4)(\alpha_1-8)(\alpha_1-10) + \tfrac{1}{30}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-4)(\alpha_1-8)\alpha_2 - \tfrac{1}{24}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-7)\alpha_3 - \tfrac{1}{2}\alpha_1(\alpha_1-1)\alpha_3-\tfrac{1}{3}\alpha_1(\alpha_1-2)(\alpha_1-4)\alpha_4 + \tfrac{1}{6}\alpha_1(\alpha_1-2)(\alpha_1-4)\alpha_2(\alpha_2-1)+$$ (see next page)

$$+\tfrac{1}{2}(\alpha_1-1)(\alpha_1-2)\alpha_2\alpha_3-\tfrac{1}{2}\alpha_1(\alpha_1-3)\alpha_5$$
$$+(\alpha_1-1)(\alpha_3-1)\alpha_3-\tfrac{1}{2}\alpha_2(\alpha_2-1)\alpha_3-\alpha_2\alpha_5+\alpha_3\alpha_4.$$

$$\begin{aligned}
D(n-8,7,1)_{(\alpha)} = {} & (1/8.6!)\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-4)(\alpha_1-5)(\alpha_1-7)(\alpha_1-1 \\
& +\tfrac{1}{144}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-5)(\alpha_1-10)\alpha_2 \\
& +\tfrac{1}{30}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-4)(\alpha_1-8)\alpha_3 \\
& +\tfrac{1}{8}\alpha_1(\alpha_1-1)(\alpha_1-3)(\alpha_1-6)\alpha_4 \\
& +\tfrac{1}{16}\alpha_1(\alpha_1-1)(\alpha_1-3)(\alpha_1-6)\alpha_2(\alpha_2-1) \\
& +\tfrac{1}{3}\alpha_1(\alpha_1-2)(\alpha_1-4)\alpha_2\alpha_3+\tfrac{1}{3}\alpha_1(\alpha_1-2)(\alpha_1-4)\alpha_5 \\
& +\tfrac{1}{2}(\alpha_1-1)(\alpha_1-2)\alpha_6+\tfrac{1}{2}(\alpha_1-1)(\alpha_1-2)\alpha_2\alpha_4 \\
& +\tfrac{1}{12}(\alpha_1-1)(\alpha_1-2)\alpha_2(\alpha_2-1)(\alpha_2-2) \\
& +\tfrac{1}{4}(\alpha_1-1)(\alpha_1-2)\alpha_3(\alpha_3-1)-(\alpha_1-1)\alpha_2\alpha_5 \\
& -\tfrac{1}{2}(\alpha_1-1)\alpha_2(\alpha_2-1)\alpha_3-(\alpha_1-1)\alpha_3\alpha_4-(\alpha_1-1)\alpha_7 \\
& -\tfrac{1}{24}\alpha_2(\alpha_2-1)(\alpha_2-2)(\alpha_2-3)-\tfrac{1}{2}\alpha_2\alpha_3(\alpha_3-1) \\
& -\tfrac{1}{2}\alpha_2(\alpha_2-1)\alpha_4-\alpha_2\alpha_6-\alpha_3\alpha_5-\tfrac{1}{2}\alpha_4(\alpha_4-1)-\alpha_8
\end{aligned}$$

$$\begin{aligned}
D(n-8,6,2)_{(\alpha)} = {} & (1/2^5.3^2.7)\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-4)(\alpha_1-5)(\alpha_1-8)(\alpha_1- \\
& +\tfrac{1}{72}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1^2-15\alpha_1+53)\alpha_2 \\
& +\tfrac{1}{24}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-5)(\alpha_1-7)\alpha_3 \\
& +\tfrac{1}{12}\alpha_1(\alpha_1-1)(\alpha_1^2-9\alpha_1+17)\alpha_2(\alpha_2-1) \\
& +\tfrac{1}{12}\alpha_1(\alpha_1-1)(\alpha_1-4)(\alpha_1-5)\alpha_4 \\
& +\tfrac{1}{6}(\alpha_1-1)(\alpha_1-2)\alpha_2(\alpha_2-1)(\alpha_2-2) \\
& +\tfrac{1}{6}\alpha_1(\alpha_1-1)(\alpha_1-5)\alpha_2\alpha_3+\tfrac{1}{2}(\alpha_1-1)\alpha_2(\alpha_2-1)\alpha_3 \\
& -\tfrac{1}{4}(\alpha_1-1)(\alpha_1-2)\alpha_3(\alpha_3-1)-(\alpha_1-1)\alpha_3\alpha_4 \\
& -\tfrac{1}{2}\alpha_1(\alpha_1-3)\alpha_6-(\alpha_1-1)\alpha_7 \\
& +\tfrac{1}{6}\alpha_2(\alpha_2-1)(\alpha_2-2)(\alpha_2-6)+\alpha_2(\alpha_2-2)\alpha_4 \\
& +\tfrac{1}{2}\alpha_2\alpha_3(\alpha_3-1)+(\alpha_2-1)\alpha_6
\end{aligned}$$

$$\begin{aligned}
D(n-8,5,3)_{(\alpha)} = {} & (1/2.6!)\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-4)(\alpha_1-5)(\alpha_1-9)(\alpha_1- \\
& +\tfrac{1}{72}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1^2-15\alpha_1+53)\alpha_2 \\
& +\tfrac{1}{12}\alpha_1(\alpha_1-1)(\alpha_1^2-9\alpha_1+17)\alpha_2(\alpha_2-1) \\
& +\tfrac{1}{120}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-9)\alpha_3 \\
& -\tfrac{1}{12}\alpha_1(\alpha_1-1)(\alpha_1-4)(\alpha_1-5)\alpha_4 \\
& +\tfrac{1}{6}\alpha_1(\alpha_1-1)(\alpha_1-5)\alpha_2\alpha_3-\tfrac{1}{3}\alpha_1(\alpha_1-2)(\alpha_1-4)\alpha_5 \\
& +\tfrac{1}{6}(\alpha_1-1)(\alpha_1-2)\alpha_2(\alpha_2-1)(\alpha_2-2) \\
& +\tfrac{1}{4}\alpha_1(\alpha_1-3)\alpha_3(\alpha_3-1)+\tfrac{1}{2}(\alpha_1-1)\alpha_2(\alpha_2-1)\alpha_3 \\
& +(\alpha_1-1)\alpha_3\alpha_4-\tfrac{1}{2}\alpha_1(\alpha_1-3)\alpha_6 \\
& -\tfrac{1}{6}\alpha_2(\alpha_2-1)(\alpha_2-2)(\alpha_2-4)-\alpha_2(\alpha_2-2)\alpha_4 \\
& +\tfrac{1}{2}\alpha_2\alpha_3(\alpha_3-1)-\alpha_2\alpha_6+\alpha_3\alpha_5
\end{aligned}$$

$$\begin{aligned}
D(n-8,4^2)_{(\alpha)} = {} & (1/5!4!)\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-4)(\alpha_1-5)(\alpha_1-10)(\alpha_1 \\
& +\tfrac{1}{360}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(2\alpha_1^2-30\alpha_1+103)\alpha_2 \\
& -\tfrac{1}{120}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1^2-12\alpha_1+47)\alpha_3-
\end{aligned}$$
(see next p

$$
\begin{aligned}
&- \tfrac{1}{12}\alpha_1(\alpha_1-1)(\alpha_1-4)(\alpha_1-5)\alpha_4 \\
&+ \tfrac{1}{24}\alpha_1(\alpha_1-1)(\alpha_1-4)(\alpha_1-5)\alpha_2(\alpha_2-1) \\
&+ \tfrac{1}{6}\alpha_1(\alpha_1^2-6\alpha_1+11)\alpha_2\alpha_3 - \tfrac{1}{6}\alpha_1(\alpha_1-1)(\alpha_1-5)\alpha_5 \\
&+ \tfrac{1}{2}\alpha_1(\alpha_1-3)\alpha_3(\alpha_3-1) - (\alpha_1-1)\alpha_2\alpha_5 \\
&- \tfrac{1}{2}(\alpha_1-1)\alpha_2(\alpha_2-1)\alpha_3 + (\alpha_1-1)\alpha_3\alpha_4 \\
&+ \alpha_2(\alpha_2-2)\alpha_4 - (\alpha_2-1)\alpha_3(\alpha_3-1) \\
&+ \tfrac{1}{4}\alpha_2(\alpha_2-1)(\alpha_2-2)(\alpha_2-5) - \alpha_3\alpha_5 + \alpha_4(\alpha_4-1).
\end{aligned}
$$

3. The characters of the four-element-partition representations $\mathcal{D}(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$. On developing, as before, the four-rowed determinant $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$ in terms of the first row we obtain

$$
\begin{aligned}
\{\lambda_1, \lambda_2, \lambda_3, \lambda_4\} &= q_{\lambda_1}(s)\{\lambda_2, \lambda_3, \lambda_4\} \\
&- q_{\lambda_1+1}(s)[\{\lambda_2, \lambda_3, \lambda_4-1\} + \{\lambda_2, \lambda_3-1, \lambda_4\} + \{\lambda_2-1, \lambda_3, \lambda_4\}] \\
&+ q_{\lambda_1+2}(s)[\{\lambda_2, \lambda_3-1, \lambda_4-1\} + \{\lambda_2-1, \lambda_3, \lambda_4-1\} + \{\lambda_2-1, \lambda_3-1, \lambda_4\}] \\
&- q_{\lambda_1+3}(s)\{\lambda_2-1, \lambda_3-1, \lambda_4-1\}.
\end{aligned}
$$

The coefficient of $-q_{\lambda_1+1}(s)$ in this expression is

$$\frac{\partial}{\partial s_1}\{\lambda_2, \lambda_3, \lambda_4\}$$

whilst that of $q_{\lambda_1+2}(s)$ is

$$\tfrac{1}{2}\left(\frac{\partial^2}{\partial s_1^{\,2}} - 2\frac{\partial}{\partial s_2}\right)\{\lambda_2, \lambda_3, \lambda_4\}.$$

Similarly the coefficient of $-q_{\lambda_1+3}(s)$ is

$$\tfrac{1}{6}\left\{\frac{\partial^2}{\partial s_1^{\,3}} - 3\frac{\partial}{\partial s_1}\left(2\frac{\partial}{\partial s_2}\right) + 2\frac{\partial}{\partial s_1}\left(3\frac{\partial}{\partial s_3}\right)\right\}\{\lambda_2, \lambda_3, \lambda_4\}$$

but it is simpler to leave it as it is. On equating the coefficients of

$$\frac{1}{\alpha_1!\cdots\alpha_n!}\left(\frac{s_1}{1}\right)^{\alpha_1}\cdots\left(\frac{s_n}{n}\right)^{\alpha_n}$$

on both sides we obtain

$$
\begin{aligned}
D(\lambda_1, \lambda_2, \lambda_3, \lambda_4)_{(\alpha)} &= \sum_{(\beta)}\{\lambda_2, \lambda_3, \lambda_4\}_{(\beta)}\binom{\alpha_1}{\beta_1}\cdots\binom{\alpha_n}{\beta_n} \\
&- \sum_{(\beta')}\{\lambda_2, \lambda_3, \lambda_4\}_{(\beta)}\binom{\alpha_1}{\beta_1'}\cdots\binom{\alpha_n}{\beta_n} \\
&+ \tfrac{1}{2}\sum_{(\beta'')}[\{\lambda_2, \lambda_3, \lambda_4\}_{\beta_1''+2} - \{\lambda_2, \lambda_3, \lambda_4\}_{\beta_2''+1}]\binom{\alpha_1}{\beta_1''}\binom{\alpha_2}{\beta_2''}\cdots\binom{\alpha_n}{\beta_n} \\
&- \sum_{(\gamma)}\{\lambda_2-1, \lambda_3-1, \lambda_4-1\}_{(\gamma)}\binom{\alpha_1}{\gamma_1}\cdots\binom{\alpha_n}{\gamma_n}.
\end{aligned}
$$

In this formula (β) is an arbitrary class of the symmetric group on $\lambda_2+\lambda_3+\lambda_4$ letters; (β') an arbitrary class of the symmetric group on $\lambda_2+\lambda_3+\lambda_4-1$ letters; (β'') an arbitrary class of the symmetric group on $\lambda_2+\lambda_3+\lambda_4-2$ letters and (γ) an arbitrary class of the symmetric group on $\lambda_2+\lambda_3+\lambda_4-3$ letters. In the second summation $\beta_1=\beta_1'+1$ and in the third summation the coefficient (in square parentheses) is the difference between the characters of $D(\lambda_2,\lambda_3,\lambda_4)$ for the classes $(\beta_1''+2,\beta_2'',\cdots,\beta_n)$ and $(\beta_1'',\beta_2''+1,\cdots,\beta_n)$ of the symmetric group on $\lambda_2+\lambda_3+\lambda_4$ letters. The first two summations may be combined in exactly the same manner as in the case of the three-element-partition representations and we thus obtain the formula:

$$
\begin{aligned}
(3)\quad & D(\lambda_1,\lambda_2,\lambda_3,\lambda_4)_{(\alpha)} \\
&= \sum_{(\beta')}\frac{1}{\beta_1!}\{\lambda_2,\lambda_3,\lambda_4\}_{(\beta)}\alpha_1(\alpha_1-1)\cdots(\alpha_1-\beta_1'+1)(\alpha_1-2\beta_1'-1)\binom{\alpha_2}{\beta_2}\cdots\binom{\alpha_n}{\beta_n} \\
&+\sum_{(\delta)}\{\lambda_2,\lambda_3,\lambda_4\}_{(\delta)}\binom{\alpha_2}{\delta_2}\cdots\binom{\alpha_n}{\delta_n} \\
&+\tfrac{1}{2}\sum_{(\beta')}[\{\lambda_2,\lambda_3,\lambda_4\}_{\beta_1''+2}-\{\lambda_2,\lambda_3,\lambda_4\}_{\beta_2''+1}]\binom{\alpha_1}{\beta_1''}\binom{\alpha_2}{\beta_2''}\cdots\binom{\alpha_n}{\beta_n} \\
&-\sum_{(\gamma)}\{\lambda_2-1,\lambda_3-1,\lambda_4-1\}_{(\gamma)}\binom{\alpha_1}{\gamma_1}\cdots\binom{\alpha_n}{\gamma_n}
\end{aligned}
$$

where (δ) is any class of the symmetric group on $\lambda_2+\lambda_3+\lambda_4$ letters *which does not contain any unary cycle* (= fixed letter).

As examples of the use of this formula we list the following:

$$
\begin{aligned}
D(n-3,1^3)_{(\alpha)} &= \tfrac{1}{6}(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)-(\alpha_1-1)\alpha_2+\alpha_3 \\
D(n-4,2,1^2)_{(\alpha)} &= \tfrac{1}{8}\alpha_1(\alpha_1-2)(\alpha_1-3)(\alpha_1-5) \\
&\quad -\tfrac{1}{2}\alpha_1(\alpha_1-3)\alpha_2-\tfrac{1}{2}\alpha_2(\alpha_2-1)+\alpha_4 \\
D(n-5,3,1^2)_{(\alpha)} &= \tfrac{1}{20}\alpha_1(\alpha_1-1)(\alpha_1-3)(\alpha_1-4)(\alpha_1-7) \\
&\quad -(\alpha_1-1)\alpha_2(\alpha_2-2)+\alpha_5 \\
D(n-5,2^2,1)_{(\alpha)} &= \tfrac{1}{24}\alpha_1(\alpha_1-1)(\alpha_1-3)(\alpha_1-5)(\alpha_1-6) \\
&\quad -\tfrac{1}{6}(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)\alpha_2 \\
&\quad -\tfrac{1}{2}\alpha_1(\alpha_1-3)\alpha_3+\tfrac{1}{2}(\alpha_1-1)\alpha_2(\alpha_2-1) \\
&\quad +(\alpha_1-1)\alpha_4-\alpha_2\alpha_3 \\
D(n-6,4,1^2)_{(\alpha)} &= \tfrac{1}{72}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-4)(\alpha_1-5)(\alpha_1-9) \\
&\quad +\tfrac{1}{12}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-7)\alpha_2+\alpha_1(\alpha_1-2)\alpha_2 \\
&\quad +\tfrac{1}{6}\alpha_1(\alpha_1-1)(\alpha_1-5)\alpha_3+(\alpha_1-1)\alpha_3 \\
&\quad -\tfrac{1}{2}\alpha_1(\alpha_1-3)\alpha_2(\alpha_2-1)-(\alpha_1-1)\alpha_2\alpha_3 \\
&\quad -\tfrac{1}{3}\alpha_2(\alpha_2-1)(\alpha_2-2)+\tfrac{1}{2}\alpha_3(\alpha_3-1)+\alpha_6
\end{aligned}
$$

$$D(n-6,3,2,1)_{(a)} = \tfrac{1}{45}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-4)(\alpha_1-6)(\alpha_1-8) - \tfrac{1}{3}\alpha_1(\alpha_1-1)(\alpha_1-5)\alpha_3 - (\alpha_1-1)\alpha_3 - \alpha_3(\alpha_3-1) + (\alpha_1-1)\alpha_5$$

$$D(n-6,2^3)_{(a)} = \tfrac{1}{144}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-5)(\alpha_1-6)(\alpha_1-7) - \tfrac{1}{24}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-7)\alpha_2 - \tfrac{1}{2}\alpha_1(\alpha_1-1)\alpha_2 - \tfrac{1}{2}\alpha_1\alpha_2(\alpha_2-3) - \tfrac{1}{6}\alpha_1(\alpha_1-1)(\alpha_1-5)\alpha_3 + \tfrac{1}{4}(\alpha_1+1)(\alpha_1-2)\alpha_2(\alpha_2-1) + \tfrac{1}{2}\alpha_1(\alpha_1-3)\alpha_4 - (\alpha_1-1)\alpha_2\alpha_3 + \tfrac{1}{2}\alpha_2(\alpha_2-1)(\alpha_2-2) - (\alpha_2-1)\alpha_4 + \alpha_3(\alpha_3-2)$$

$$D(n-7,5,1^2)_{(a)} = \tfrac{1}{336}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-5)(\alpha_1-6)(\alpha_1-11) + \tfrac{1}{24}\alpha_1(\alpha_1-1)(\alpha_1-3)(\alpha_1-5)(\alpha_1-6)\alpha_2 + \tfrac{1}{8}\alpha_1(\alpha_1-2)(\alpha_1-3)(\alpha_1-5)\alpha_3 + \tfrac{1}{6}(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)\alpha_4 - \tfrac{1}{12}(\alpha_1+1)(\alpha_1-1)(\alpha_1-6)\alpha_2(\alpha_2-1) - \tfrac{1}{2}\alpha_1(\alpha_1-3)\alpha_2\alpha_3 - (\alpha_1-1)\alpha_2\alpha_4 - \tfrac{1}{2}(\alpha_1-1)\alpha_2(\alpha_2-1)(\alpha_2-2) - \tfrac{1}{2}\alpha_2(\alpha_2-1)\alpha_3 + \alpha_3\alpha_4 + \alpha_7$$

$$D(n-7,4,2,1)_{(a)} = \tfrac{1}{144}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-5)(\alpha_1-7)(\alpha_1-10) + \tfrac{1}{24}\alpha_1(\alpha_1-1)(\alpha_1-3)(\alpha_1-5)(\alpha_1-6)\alpha_2 - \tfrac{1}{24}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-7)\alpha_3 - \tfrac{1}{6}(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)\alpha_4 - \tfrac{1}{12}(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)\alpha_2(\alpha_2-1) - \tfrac{1}{2}\alpha_1(\alpha_1-3)\alpha_2\alpha_3 + (\alpha_1-1)\alpha_2\alpha_4 + \tfrac{1}{6}(\alpha_1-1)\alpha_2(\alpha_2-1)(\alpha_2-2) - \tfrac{1}{2}(\alpha_1-1)\alpha_3(\alpha_3-1) + (\alpha_1-1)\alpha_6 - \tfrac{1}{2}\alpha_2(\alpha_2-1)\alpha_3 - \alpha_3\alpha_4$$

$$D(n-7,3^2,1)_{(a)} = \tfrac{1}{240}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-5)(\alpha_1-8)(\alpha_1-9) + \tfrac{1}{120}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-9)\alpha_2 - \tfrac{1}{8}\alpha_1(\alpha_1-2)(\alpha_1-3)(\alpha_1-5)\alpha_3 - \tfrac{1}{6}\alpha_1(\alpha_1-1)(\alpha_1-5)\alpha_4 + \tfrac{1}{12}(\alpha_1^2-5\alpha_1+12)(\alpha_1-1)\alpha_2(\alpha_2-1) + \tfrac{1}{2}\alpha_1(\alpha_1-3)\alpha_2\alpha_3 + \tfrac{1}{2}\alpha_1(\alpha_1-3)\alpha_5 - (\alpha_1-1)\alpha_2\alpha_4 - \tfrac{1}{2}(\alpha_1-1)\alpha_2(\alpha_2-1)(\alpha_2-2) + \alpha_2\alpha_5 + \tfrac{1}{2}\alpha_2(\alpha_2-1)\alpha_3 - \alpha_3\alpha_4$$

$$D(n-7,3,2^2)_{(a)} = \tfrac{1}{240}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-6)(\alpha_1-7)(\alpha_1-9) - \tfrac{1}{120}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-9)\alpha_2 - \tfrac{1}{8}\alpha_1(\alpha_1-1)(\alpha_1-3)(\alpha_1-6)\alpha_3 + \text{(see next page)}$$

$$
\begin{aligned}
&+ \tfrac{1}{6}(\alpha_1 - 1)(\alpha_1 - 2)(\alpha_1 - 3)\alpha_4 \\
&+ \tfrac{1}{12}\alpha_1(\alpha_1 - 1)(\alpha_1 - 5)\alpha_2(\alpha_2 - 3) \\
&- \tfrac{1}{2}(\alpha_1 - 1)(\alpha_1 - 2)\alpha_2\alpha_3 \\
&+ \tfrac{1}{2}(\alpha_1 - 1)(\alpha_1 - 2)\alpha_5 - (\alpha_1 - 1)\alpha_2\alpha_4 \\
&+ \tfrac{1}{2}(\alpha_1 - 1)\alpha_2(\alpha_2 - 1)(\alpha_2 - 3) \\
&+ \tfrac{1}{2}\alpha_2(\alpha_2 - 1)\alpha_3 - \alpha_2\alpha_5 + \alpha_3\alpha_4.
\end{aligned}
$$

$$
\begin{aligned}
D(n-8, 6, 1^2)_{(a)} = {} & (1/2^4.5\,!)\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-4)(\alpha_1-6)(\alpha_1-7)(\alpha_1-1 \\
&+ \tfrac{1}{240}\alpha_1(\alpha_1 - 1)(\alpha_1 - 2)(\alpha_1 - 4)(3\alpha_1^2 - 42\alpha_1 + 139)\alpha_2 \\
&+ \tfrac{1}{20}\alpha_1(\alpha_1 - 1)(\alpha_1 - 3)(\alpha_1 - 4)(\alpha_1 - 7)\alpha_3 \\
&+ \tfrac{1}{48}\alpha_1(\alpha_1 - 2)(\alpha_1^2 - 8\alpha_1 + 31)\alpha_2(\alpha_2 - 1) \\
&+ (\alpha_1 - 1)\alpha_2\alpha_3 - \tfrac{1}{4}\alpha_1(\alpha_1 - 3)\alpha_2(\alpha_2 - 1)(\alpha_2 - 2) \\
&- (\alpha_1 - 1)\alpha_2(\alpha_2 - 1)\alpha_3 - \tfrac{1}{2}\alpha_1(\alpha_1 - 3)\alpha_2\alpha_4 \\
&- \tfrac{1}{8}\alpha_2(\alpha_2 - 1)(\alpha_2 - 2)(\alpha_2 - 3) - (\alpha_1 - 1)\alpha_2\alpha_5 \\
&- \tfrac{1}{2}\alpha_2(\alpha_2 - 1)\alpha_4 + \tfrac{1}{8}\alpha_1(\alpha_1 - 2)(\alpha_1 - 3)(\alpha_1 - 5)\alpha_4 \\
&+ \tfrac{1}{6}(\alpha_1 - 1)(\alpha_1 - 2)(\alpha_1 - 3)\alpha_5 + \alpha_3\alpha_5 \\
&+ \tfrac{1}{2}\alpha_4(\alpha_4 - 1) + \alpha_8
\end{aligned}
$$

$$
\begin{aligned}
D(n-8,5,2,1)_{(a)} = {} & \tfrac{1}{630}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-4)(\alpha_1-6)(\alpha_1-8)(\alpha_1-12) \\
&+ \tfrac{1}{45}\alpha_1(\alpha_1 - 1)(\alpha_1 - 2)(\alpha_1 - 4)(\alpha_1 - 6)(\alpha_1 - 8)\alpha_2 \\
&+ \tfrac{1}{30}\alpha_1(\alpha_1 - 1)(\alpha_1 - 3)(\alpha_1^2 - 11\alpha_1 + 33)\alpha_3 \\
&- \tfrac{1}{3}(\alpha_1 - 1)(\alpha_1^2 - 5\alpha_1 + 3)\alpha_2\alpha_3 \\
&- \tfrac{1}{2}\alpha_1(\alpha_1 - 3)\alpha_3(\alpha_3 - 1) \\
&- \tfrac{1}{6}(\alpha_1 - 1)(\alpha_1 - 2)(\alpha_1 - 3)\alpha_5 + (\alpha_1 - 1)\alpha_2\alpha_5 \\
&+ (\alpha_1 - 1)\alpha_7 - \alpha_2\alpha_3(\alpha_3 - 1) - \alpha_3\alpha_5
\end{aligned}
$$

$$
\begin{aligned}
D(n-8,4,3,1)_{(a)} = {} & \tfrac{1}{576}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-4)(\alpha_1-6)(\alpha_1-9)(\alpha_1-11) \\
&+ \tfrac{1}{72}\alpha_1(\alpha_1 - 1)(\alpha_1 - 2)(\alpha_1 - 4)(\alpha_1 - 5)(\alpha_1 - 9)\alpha_2 \\
&- \tfrac{1}{24}\alpha_1(\alpha_1 - 1)(\alpha_1 - 3)(\alpha_1 - 5)(\alpha_1 - 6)\alpha_3 \\
&+ \tfrac{1}{24}\alpha_1(\alpha_1 - 2)(\alpha_1^2 - 8\alpha_1 + 19)\alpha_2(\alpha_2 - 1) \\
&- \tfrac{1}{6}\alpha_1(\alpha_1 - 3)\alpha_2(\alpha_2 - 1)(\alpha_2 - 2) \\
&+ \tfrac{1}{6}(\alpha_1 - 1)(\alpha_1 - 2)(\alpha_1 - 3)\alpha_2\alpha_3 \\
&- \tfrac{1}{2}(\alpha_1 - 1)\alpha_2(\alpha_2 - 1)\alpha_3 + \tfrac{1}{4}\alpha_1(\alpha_1 - 3)\alpha_3(\alpha_3 - 1) \\
&- \tfrac{1}{6}\alpha_1(\alpha_1 - 2)(\alpha_1^2 - 8\alpha_1 + 13)\alpha_4 + \tfrac{1}{2}\alpha_1(\alpha_1 - 3)\alpha_6 \\
&- (\alpha_1 - 1)\alpha_3\alpha_4 - \tfrac{1}{12}\alpha_2(\alpha_2 - 1)(\alpha_2 - 2)(\alpha_2 - 3) \\
&+ \tfrac{1}{2}\alpha_2\alpha_3(\alpha_3 - 1) + \alpha_2\alpha_6 - \alpha_4(\alpha_4 - 1)
\end{aligned}
$$

$$
\begin{aligned}
D(n-8, 4, 2^2)_{(a)} = {} & (1/6\,!)\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-4)(\alpha_1-7)(\alpha_1-8)(\alpha_1-11 \\
&+ \tfrac{1}{180}\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1^2-15\alpha_1+59)\alpha_2 \\
&- \tfrac{1}{30}\alpha_1(\alpha_1 - 1)(\alpha_1 - 2)(\alpha_1 - 4)(\alpha_1 - 8)\alpha_3 \\
&+ \tfrac{1}{3}(\alpha_1^2 - 3\alpha_1 - 1)\alpha_2(\alpha_2 - 1)(\alpha_2 - 2) \\
&- \tfrac{1}{3}\alpha_1(\alpha_1-2)(\alpha_1-4)\alpha_2\alpha_3 - \tfrac{1}{2}\alpha_1(\alpha_1-3)\alpha_2(\alpha_2-1) \\
&- \tfrac{1}{4}(\alpha_1 - 1)(\alpha_1 - 2)\alpha_3(\alpha_3 - 1) + \text{(see next page)}
\end{aligned}
$$

$$
\begin{aligned}
&+ \tfrac{1}{6}(\alpha_1 - 1)(\alpha_1 - 2)(\alpha_1 - 3)\alpha_5 \\
&+ \tfrac{1}{2}(\alpha_1 - 1)(\alpha_1 - 2)\alpha_6 - (\alpha_1 - 1)\alpha_2\alpha_5 \\
&+ \tfrac{1}{3}\alpha_2(\alpha_2 - 1)(\alpha_2 - 2)(\alpha_2 - 3) \\
&+ \tfrac{1}{2}\alpha_2\alpha_3(\alpha_3 - 1) + \alpha_3\alpha_5 - \alpha_2\alpha_6
\end{aligned}
$$

$$
\begin{aligned}
D(n-8, 3^2, 2)_{(\alpha)} = {} & (1/8.5\,!)\alpha_1(\alpha_1-1)(\alpha_1-2)(\alpha_1-3)(\alpha_1-4)(\alpha_1-7)(\alpha_1-9)(\alpha_1-10) \\
&- \tfrac{1}{24}\alpha_1(\alpha_1 - 1)(\alpha_1 - 2)(\alpha_1 - 7)\alpha_2 \\
&- \tfrac{1}{20}\alpha_1(\alpha_1 - 1)(\alpha_1 - 3)(\alpha_1 - 4)(\alpha_1 - 7)\alpha_3 \\
&+ \tfrac{1}{24}\alpha_1(\alpha_1 - 1)(\alpha_1 - 4)(\alpha_1 - 5)\alpha_2(\alpha_2 - 1) \\
&+ \tfrac{1}{3}\alpha_1(\alpha_1 - 2)(\alpha_1 - 4)\alpha_5 - (\alpha_1^2 - 3\alpha_1 + 1)\alpha_2\alpha_4 \\
&+ \tfrac{1}{2}\alpha_1(\alpha_1 - 3)\alpha_4 \\
&- \tfrac{1}{2}\alpha_1\alpha_2(\alpha_2 - 1) + (\alpha_1 - 1)\alpha_2(\alpha_2 - 2)\alpha_3 \\
&- \tfrac{1}{4}\alpha_2(\alpha_2-1)(\alpha_2-2)(\alpha_2-5) - \alpha_3\alpha_5 + \alpha_4(\alpha_4-1).
\end{aligned}
$$

4. Concluding remarks. Whilst the expressions given in the preceding paragraphs are somewhat lengthy their evaluation for any given class (α) of the symmetric group on n letters is quite easy since a great many terms of each expression vanish for any given class (α). The formulae explicitly written out suffice to give all characters for $n \leqq 8$; for $n = 9$ they give all save the characters of the self-associated five-element-partition $(5, 1^4)$; it being always understood that of two associated partitions of n we apply our formula to the one whose partition contains the smaller number of elements. For $n = 10$ our formulae yield all characters save those corresponding to the partitions $(6, 1^4)$, $(5, 2, 1^3)$ the second of which is self associated. For $n = 11$ the only partitions not cared for are $(7, 1^4)$, $(6, 2, 1^3)$, $(6, 1^5)$, $(5, 3, 1^3)$ whilst for $n = 12$ they are $(8, 1^4)$, $(7, 2, 1^3)$, $(7, 1^5)$, $(6, 3, 1^3)$, $(6, 2^2, 1^2)$, $(6, 2, 1^4)$, $(5, 4, 1^3)$, $(5, 3, 2, 1^2)$. Thus if we are concerned only with partitions containing not more than four elements the characters are all cared for, up to $n = 12$, by our formulae. The missing partitions, up to $n = 12$, may be cared for since they all end in 1 by constructing formulae as described in the following paragraph.

Since

$$
\{\lambda_1, \cdots, \lambda_k\} \cdot \{1\} = \{\lambda_1, \cdots, \lambda_k, 1\} + \{\lambda_1 + 1, \cdots, \lambda_k\} + \cdots + \{\lambda_1, \cdots, \lambda_k + 1\}
$$

(p. 480) we find at once, on comparing the coefficients of

$$
\frac{1}{\alpha_1! \cdots \alpha_n!}\left(\frac{s_1}{1}\right)^{\alpha_1} \cdots \left(\frac{s_n}{n}\right)^{\alpha_n}
$$

(where (α) is an arbitrary class of the symmetric group on $\lambda_1 + \cdots + \lambda_k + 1$ letters) on both sides the formula:

$$(4) \qquad D(\lambda_1, \cdots, \lambda_k, 1) = \alpha_1 D(\lambda_1, \cdots, \lambda_k)_{(\alpha')} \\ - [D(\lambda_1 + 1, \cdots, \lambda_k)_{(\alpha)} + \cdots + D(\lambda_1, \cdots, \lambda_k + 1)_{(\alpha)}]$$

where $\alpha'_1 = \alpha_1 - 1$ so that $(\alpha') = (\alpha_1 - 1, \alpha_2, \cdots, \alpha_n)$ is a class of the symmetric group on $\lambda_1 + \lambda_2 + \cdots + \lambda_k$ letters. As an instance of the application of this formula consider $D(n-4, 1^4)$: here

$$\lambda_1 = n - 4, \quad \lambda_2 = \lambda_3 = \lambda_4 = 1$$

and so

$$D(n-4, 1^4)_{(\alpha)} = \alpha_1 D(n-4, 1^3)_{(\alpha')} \\ - [D(n-3, 1^3)_{(\alpha)} + D(n-4, 2, 1^2)_{(\alpha)}].$$

Since n does not enter explicitly the previously derived formula for $D(n-3, 1^3)_{(\alpha)}$ we read off the expression for $D(n-4, 1^3)_{(\alpha')}$ by merely replacing α_1 by $\alpha_1 - 1$ in the expression for $D(n-3, 1^3)_{(\alpha)}$. In this way we find

$$\begin{aligned} D(n-4, 1^4)_{(\alpha)} &= \tfrac{1}{24}(\alpha_1 - 1)(\alpha_1 - 2)(\alpha_1 - 3)(\alpha_1 - 4) \\ &\quad - \tfrac{1}{2}(\alpha_1 - 1)(\alpha_1 - 2)\alpha_2 + (\alpha_1 - 1)\alpha_3 \\ &\quad + \tfrac{1}{2}\alpha_2(\alpha_2 - 1) - \alpha_4 \end{aligned}$$

Similarly

$$\begin{aligned} D(n-5, 2, 1^3)_{(\alpha)} &= \tfrac{1}{30}\alpha_1(\alpha_1 - 2)(\alpha_1 - 3)(\alpha_1 - 4)(\alpha_1 - 6) \\ &\quad - \tfrac{1}{3}\alpha_1(\alpha_1 - 2)(\alpha_1 - 4)\alpha_2 \\ &\quad + \tfrac{1}{2}\alpha_1(\alpha_1 - 3)\alpha_3 + \alpha_2\alpha_3 - \alpha_5 \\ D(n-5, 1^5)_{(\alpha)} &= \tfrac{1}{120}(\alpha_1 - 1)(\alpha_1 - 2)(\alpha_1 - 3)(\alpha_1 - 4)(\alpha_1 - 5) \\ &\quad - \tfrac{1}{6}(\alpha_1 - 1)(\alpha_1 - 2)(\alpha_1 - 3)\alpha_2 \\ &\quad + \tfrac{1}{2}(\alpha_1 - 1)(\alpha_1 - 2)\alpha_3 \\ &\quad + \tfrac{1}{2}(\alpha_1 - 1)\alpha_2(\alpha_2 - 1) - (\alpha_1 - 1)\alpha_4 - \alpha_2\alpha_3 + \alpha_5 \\ D(n-6, 3, 1^3)_{(\alpha)} &= \tfrac{1}{72}\alpha_1(\alpha_1 - 1)(\alpha_1 - 3)(\alpha_1 - 4)(\alpha_1 - 5)(\alpha_1 - 8) \\ &\quad - \tfrac{1}{12}(\alpha_1 - 1)(\alpha_1 - 2)(\alpha_1 - 3)(\alpha_1 - 4)\alpha_2 \\ &\quad + (\alpha_1 - 1)(\alpha_1 - 3)\alpha_2 \\ &\quad - \tfrac{1}{2}(\alpha_1 - 1)(\alpha_1 - 2)\alpha_2(\alpha_2 - 1) \\ &\quad + \tfrac{1}{3}\alpha_2(\alpha_2 - 1)(\alpha_2 - 2) \\ &\quad + \tfrac{1}{6}\alpha_1(\alpha_1 - 1)(\alpha_1 - 5)\alpha_3 + (\alpha_1 - 1)\alpha_2\alpha_3 \\ &\quad + \tfrac{1}{2}\alpha_3(\alpha_3 - 1) - \alpha_6 \end{aligned}$$

These formulae, together with those already given, furnish all the characters of the various symmetric groups for $n \leq 11$. The characters of the symmetric groups up to $n = 13$ have recently been published (4) and the provision of the formulae of the present paper would be without point if they

merely enabled us to provide data already known by other methods. The whole point of the present investigation is that the formulae given are available for *any value* of n. To be explicit suppose we wish to know the characters of the symmetric group on $n = 16$ letters. This group has 231 irreducible representations and the mere printing of the character table presents difficulties. Of the 231 representations some 88 $(= 44 \times 2)$ have their characters given by the formulae of the present paper. In addition to these we have the identity representation and its associate, the alternating representation, so that we know, without recurrence methods, the characters of 90 out of the 231 irreducible representations. Of the 231 partitions of 16, sixty-four contain not more than four elements and our formulae furnish the characters of 41 of the 64 corresponding irreducible representations.

THE JOHNS HOPKINS UNIVERSITY,
INSTITUTE FOR ADVANCED STUDY.

REFERENCES.

1. F. D. Murnaghan, "On the representations of the symmetric group," *American Journal of Mathematics,* **59** (1937), pp. 437-488.
2. G. Frobenius, "Über die Charaktere der symmetrischen Gruppe," *Berliner Berichte* (1900), pp. 516-534.
3. D. E. Littlewood and A. R. Richardson, "Some special S-functions and q-series," *Quarterly Journal of Mathematics,* Oxford Ser. **6** (1935), pp. 184-198.
4. M. Zia-ud-Din, "The characters of the symmetric group of degrees 12 and 13," *Proceedings of the London Mathematical Society* (2), **42** (1937), pp. 340-355.

ISOTROPIC STATIC SOLUTIONS OF THE FIELD EQUATIONS IN EINSTEIN'S THEORY OF GRAVITATION.*

By P. Y. Chou.

1. Introduction. In the classic Schwarzschild's solution of the gravitational field of a mass particle it is well known that a system of isotropic coördinates exists in which the velocity of light is the same in every direction at any point exterior to the particle. The question naturally arises whether there are other isotropic solutions of Einstein's field equations, and if they exist, we should like to know the kind of distribution of matter that produces the corresponding gravitational fields.

We shall limit our investigation to static fields only and shall give a detailed proof of the necessary and sufficient condition for the existence of isotropic fields in empty space, while proof for the corresponding theorem within matter will only be indicated in general outlines. We shall see that for a type of problem involving the determination of the field of a single body so that space outside the body is free from other singularities, there are two classes of solutions and as special cases of each class we find respectively Schwarzschild's solution and the solution for a semi-infinite plane with variable distribution of mass which was not known before. As a third example of our general result we give Kasner's solution which forms a class by itself. From the field equations within matter in general we can also prove incidentally that the Einstein static universe is the only solution for a closed static space filled with matter which is kept at constant pressure everywhere without assuming the spherical symmmetry property of the universe to start with.

2. Isotropic static gravitational fields in empty space. The arc element of a static four dimensional continuum takes the form

$$ds^2 = U^2 dt^2 + g_{ij} dx^i dx^j, \qquad (i, j = 1, 2, 3) \tag{2.1}$$

* Received April 23, 1937.

where $t = x^0$ denotes the time-like and x^1, x^2, x^3, the space-like coördinates and the functions U, g_{ij} are functions of x^1, x^2, x^3 only. We use Latin letters to denote the indices, 1, 2, 3, of the space variables, while Greek letters can take four indices including the time coördinate x^0. The ten field equations with the cosmological constant absent,

$$G_{\alpha\beta} = 0 \tag{2.2}$$

are split in the present case into

$$G_{ij} = R_{ij} + \frac{1}{U} U_{i,j} = 0, \qquad G_{0i} = 0, \qquad G_{00} = U U^i{}_{,i} = 0, \tag{2.3}$$

where R_{ij} is the contracted Riemann-Christoffel tensor for the 3-space, $t = \text{const.}$, of the arc-element (2.1), and we denote, for brevity,

$$U_i = \frac{\partial U}{\partial x^i}, \qquad U^i = g^{ij} U_j;$$

$$U_{i,j} = \frac{\partial U_i}{\partial x^j} - \left\{ \begin{matrix} h \\ ij \end{matrix} \right\} U_h, \qquad U^i{}_{,i} = g^{ij} U_{i,j},$$

covariant differentiation being taken with respect to the metric tensor g_{ij}. Since we only deal with transformations of the space-like coördinates, U is an invariant, U_i and its higher covariant derivatives are all tensors.

We assume that within the regions of the 3-space under investigation the functions U and g_{ij} possess continuous derivatives to any degree we need. Furthermore we are only interested in the real solutions of (2.3) for these gravitational potentials and real transformations of coördinates so that two solutions which can not be transformed into each other by a real transformation are considered as distinct. Next we shall establish the validity of the following theorem:

THEOREM. *A necessary and sufficient condition for the static field equations* (2.3) *to possess isotropic solutions in which the arc-element* (2.1) *can be transformed into the canonical form*

$$ds^2 = U^2 dt^2 - e^{-2\sigma}(dx^2 + dy^2 + dz^2) \tag{2.4}$$

is that the function U should satisfy the following set of differential equations,

$$(2.5) \qquad U_{i,j} = \frac{6U}{c+U^2}\left(U_iU_j - \tfrac{1}{3}g_{ij}U^hU_h\right), \quad \text{and} \quad R = 0,$$

where R is the scalar curvature invariant for the 3-space, $t = const.$, and c is an arbitrary constant.

If we contract (2. 5) with U^j, the resultant equations are integrable giving

$$(2.6) \qquad U^hU_h = -k^2(c+U^2)^4$$

where k^2 is another constant of integration.

We begin to prove the theorem by establishing the necessary condition first. If the arc-element of the 3-space in (2. 1) is conformal to a flat space as in (2. 4), then the contracted Riemann-Christoffel tensor R_{ij} satisfies the necessary condition [1]

$$(2.7) \qquad R_{ij,k} - R_{ik,j} - \tfrac{1}{4}(g_{ij}R_{,k} - g_{ik}R_{,j}) = 0$$

where $R_{ij,k}$ denotes as usual the covariant derivative of R_{ij} with respect to the metric tensor g_{ij} in (2. 1). From (2. 3) we have $R = 0$ and

$$(2.8) \qquad U^hR_{hijk} - \frac{1}{U}(U_{i,j}U_k - U_{i,k}U_j) = 0$$

in which we have utilized the well-known relation

$$(2.9) \qquad U_{i,jk} - U_{i,kj} = U^hR_{hijk}.$$

Now for any 3-space Weyl's conformal tensor vanishes identically or the Riemann-Christoffel tensor is expressible as [2]

$$(2.10) \quad R_{hijk} = -g_{jh}R_{ik} + g_{hk}R_{ij} - g_{ik}R_{hj} + g_{ij}R_{hk} - \frac{R}{2}(g_{hk}g_{ij} - g_{hj}g_{ik}).$$

Combining (2. 3) with (2. 8) and (2. 10) we obtain the following relation on the covariant derivatives of U

$$(2.11) \qquad 2(U_{i,j}U_k - U_{i,k}U_j) + g_{ij}U^hU_{h,k} - g_{ik}U^hU_{h,j} = 0.$$

---

[1] L. P. Eisenhart, *Riemannian Geometry* (1926), which will be referred to as "R. G.", p. 92.

[2] R. G., p. 90.

To solve (2. 11) for U_{ij} in terms of U and its first derivatives we contract (2. 11) by U^k and U^i in succession and eliminate the intermediate expression, $U^k U_{k;j}$. The result can be written

$$U_{i,j} = \alpha(U_i U_j - \tfrac{1}{3} g_{ij} U^h U_h) \tag{2. 12}$$

where α, an invariant, is defined as $\alpha = 3U^h U_{h,k} U^k / 2(U_l U^l)^2$.

This function α can be determined in the following way. We regard (2. 12) as a set of equations defining the function $U_{i,j}$ which, on the other hand, should satisfy the condition of integrability (2. 9). Putting (2. 12) in (2. 9), then contracting with U^i and meanwhile utilizing the relations (2. 10) and the field equations (2. 3), we find that α must be a function of U alone and furthermore its derivative with respect to U is given by

$$\frac{d\alpha}{dU} = \left(\frac{\alpha}{U} - \frac{\alpha^2}{3}\right) \tag{2. 13}$$

Integrating, we get

$$\alpha = 6U/(c + U^2) \tag{2. 14}$$

where c is a constant of integration. Hence (2. 5) as a necessary condition for the existence of isotropic solutions of the field equations (2. 3) is established.

We shall show conversely that if this condition holds true, then (2. 3) are satisfied and from the way of deriving (2. 5), condition (2. 7) is satisfied which is also sufficient for the 3-space in (2. 1) to be conformal to a flat space.

Differentiate (2. 5) covariantly with respect to g_{ij} and form the left-hand side of (2. 9). Inserting the expression for R_{hijk} in the right-hand side from (2. 10), we find

$$\begin{aligned} \frac{2}{c + U^2} U^h U_h (g_{ij} U_k - g_{ik} U_j) \\ = U_j R_{ik} - U_k R_{ij} + g_{ik} U^h R_{hj} - g_{ij} U^h R_{hk} + \frac{R}{2} (g_{ij} U_k - g_{ik} U_j). \end{aligned} \tag{2. 15}$$

Contract the indices i and k by g^{ik} and obtain

$$U^k R_{kj} = -\frac{4}{c + U^2} U^h U_h U_j. \tag{2. 16}$$

Contracting (2. 15) with U^k and eliminating $U^k R_{kj}$ by (2. 16) we get finally

$$(2.17) \quad R_{ij} = -\frac{6}{c+U^2}\,(U_iU_j - \tfrac{1}{3}g_{ij}U^hU_h) + \frac{R}{2}\left(g_{ij} - \frac{U_iU_j}{U^hU_h}\right).$$

The function R in (2.17) can be set equal to zero as follows: By utilizing the Bianchi relation which holds for any Riemannian space,[3]

$$(2.18) \qquad R^i{}_{j,i} - \tfrac{1}{2}R_{,j} = 0,$$

we can form out of (2.5), (2.6) and (2.17) a differential equation in R which can be integrated. If we choose the constant of integration properly R vanishes identically. The fact that R can be different from zero points out the possibility of existence of isotropic solutions of the field equations within matter as we shall see below.

3. Field of a single body; Kasner and Schwarzschild's solutions and field of a semi-infinite plane. We have seen in the last section that the existence and finding of the isotropic gravitational fields are equivalent to the solution of the equations (2.5). For the general problem which may involve a number of bodies, the solutions of (2.5) are very difficult to ascertain. But for the type of problem which contains only one body, then the space outside the body is free from other singularities and we can make a transformation of coördinates such that in this new system of reference the function U becomes a function of the coördinate u only. Moreover we limit ourselves to the consideration of the class of problems in which with the coördinate u thus specified the spatial part of the arc-element (2.1) can still assume the orthogonal form. Hence (2.1) can be written as

$$(3.1) \qquad ds^2 = U^2(u)\,dt^2 + g_{11}du^2 + g_{22}dv^2 + g_{33}dw^2,$$

where g_{11}, g_{22}, g_{33} are unknown functions of u, v, w to be determined.

The equations (2.5) can be computed for the quadratic form (3.1). They are immediately integrable giving

$$\frac{dU}{du} = k(c+U^2)^2(-g_{11})^{1/2},$$

$$(3.2) \quad g_{11} = g_1(u), \quad g_{22} = g_2(v,w)\,(c+U^2)^{-2}, \quad g_{33} = g_3(v,w)\,(c+U^2)^{-2},$$

[3] R. G., p. 82.

where k is a constant of integration and g_1, g_2, g_3 are arbitrary functions of their arguments indicated. We notice that equations (3.2) also satisfy (2.6) as they should.

Since any 2-space is conformal to the flat 2-space and g_{11} is a function of u only, we may write (3.1) without any loss of generality in the following form

$$(3.3) \qquad ds^2 = U^2(u)\,dt^2 - \frac{1}{(c+U^2)^2}\left[du^2 + e^{\chi(v,w)}(dv^2+dw^2)\right],$$

where

$$\frac{dU}{du} = k(c+U^2).$$

So far only the equations (2.5) involving U are satisfied. In order to satisfy every member of the field equations (2.3), the scalar curvature invariant R of the 3-space $t = \text{const.}$ in (3.3) must vanish. This condition leads to the following differential equation in χ [4]

$$(3.4) \qquad \frac{\partial^2\chi}{\partial v^2} + \frac{\partial^2\chi}{\partial w^2} - 8k^2ce^{\chi} = 0.$$

The constant c may be either negative, zero, or positive and can be normalized to be -1, 0, $+1$ respectively for simplicity. The three values of c correspond to three distinct classes of solutions. We shall show first that for the class $c = 0$, there is only one solution in the class which was discovered by Kasner and all other members of the class are transformable into Kasner's solution.

When $c = 0$, (3.4) becomes Laplace's equation for χ in the vw-plane. If we take the solution $\chi = 0$, then (3.3) becomes Kasner's solution [5]

$$(3.5) \qquad ds^2 = (ku)^{-2}dt^2 - k^4u^4(du^2+dv^2+dw^2)$$

which can be considered as describing the field of an infinite plane of constant surface density coinciding with the vw-plane. But this is also the only solution of the class. For if we take the general real solution of χ to be

[4] The computation of $R = 0$ can be accomplished by setting $T_4{}^4 = 0$ on p. 256 in Tolman's *Relativity, Thermodynamics, and Cosmology* (1934).

[5] E. Kasner, *Transactions of the American Mathematical Society*, vol. 27 (1925), p. 160.

$$\chi = f(v+iw) + f(v-iw) = f + f^*$$

where f is an arbitrary analytic function of the complex variable $z = v + iw$, then

$$e^{\chi}(dv^2 + dw^2) = e^f dz e^{f^*} dz^* = dZ dZ^* = dV^2 + dW^2,$$

where

$$dZ = e^f dz = dV + idW.$$

For the other two classes of solutions $c = \pm 1$, there is an infinite number in each class. This is due to the fact that to each solution of the partial differential equation (3.4) we can make a transformation of the coördinates from the (u, v, w)-system to the canonical system (x, y, z) of (2.4). In these transformations all of the three variables u, v, w are involved in general and the final functions $\sigma(x, y, z)$ and $U(x, y, z)$ are necessarily distinct if we start with the distinct function χ. The converse argument also holds. If two solutions $U_1(x, y, z)$, $\sigma_1(x, y, z)$ and $U_2(x, y, z)$, $\sigma_2(x, y, z)$ in the (x, y, z)-system of (2.4) can be transformed into each other by coördinate transformation, then in the (u, v, w)-system they should have the same function $\chi(v, w)$.

Due to the non-linear character of the equation (3.4), its general solution is difficult to obtain. But we can give two special solutions, one in each class. Consider the case $c = -1$. Then from (3.3) we can take

$$U^2 = \tan h^2 ku. \tag{3.6}$$

As a special solution of (3.4) we assume χ to be a function of v only. Moreover from this ordinary differential equation in v we choose the solution

$$e^{\chi} = \sin^2\theta/4k^2, \quad \text{where} \quad dv = d\theta/\sin\theta. \tag{3.7}$$

If we put $e^{-2ku} = m/2r$, $k^2 = 1/16m^2$, the arc-element (3.3) is recognizable as Schwarzschild's solution in isotropic coördinates.

As a solution for the second class $c = +1$, we also take $\chi = \chi(v)$ and the function

$$e^{\chi} = 1/4k^2v^2. \tag{3.8}$$

Take $U^2 = \cot^2 ku$ and by normalizing the coördinates such that $2ku = \phi$, $2kv = \rho$, and $2kw = z$, we find

$$ds^2 = \cot^2(\phi/2)\,dt^2 - \frac{\sin^4\phi/2}{(2k)^2\rho^2}(d\rho^2 + \rho^2 d\phi^2 + dz^2). \tag{3.9}$$

The geometrical significance of the coördinates ρ, ϕ, z is self-evident. The gravitational potentials in (3.9) are even functions of ϕ periodic in ϕ with the period 2π, their only singularity being on the positive x-axis $\phi = 0$. If we compare (3.9) and (3.5), we notice that at points where ϕ is small and ρ large the field defined by (3.9) is very much the same as that of Kasner's solution, (3.5). Hence we may interpret (3.9) as the gravitational field of a semi-infinite plane coinciding with the positive half of the zx-plane with variable distribution of mass upon the plane.

4. Isotropic fields within matter. The static field equations for the arc-element (2.1) within matter are

$$
\begin{aligned}
R_{ij} &= -\frac{U}{1} U_{i,j} + 4\pi\left(\rho_{00} - p_0 + \frac{\Lambda}{4\pi}\right) g_{ij}, \\
U^i{}_{,i} &= -4\pi\left(\rho_{00} + 3p_0 - \frac{\Lambda}{4\pi}\right) U, \qquad (4.1)\\
R &= 16\pi\left(\rho_{00} + \frac{\Lambda}{8\pi}\right),
\end{aligned}
$$

where Λ is the cosmological constant, ρ_{00} the proper density and p_0 the pressure, ρ_{00} and p_0 being invariant functions of the space-like coördinates and measured in relativistic units.

To the field equations (4.1) we can add the dynamical equations of motion,

$$P^{\mu\nu}{}_{,\nu} = 0; \qquad P^{\mu\nu} = (\rho_{00} + p_0) v^\mu v^\nu - p_0 g^{\mu\nu}.$$

In the present static case since $v^i = 0$, $v^0 = 1/U$, we find

$$P^{ij} = -p_0 g^{ij}, \qquad P^{i0} = 0, \qquad P^{00} = \rho_{00}/U^2$$

and the equations of motion degenerate into

$$p_{0,i} = -\frac{1}{U}(\rho_{00} + p_0) U_i \qquad (4.2)$$

which can also be derived by inserting (4.1) in the Bianchi identity (2.18).

From (4.1) and (4.2) we can immediately draw the conclusion that if matter is kept at a constant non-negative pressure everywhere in a closed space ($\Lambda > 0$), then the only solution of (4.1) is the Einstein static universe of relativistic cosmology. For in such a space U must be constant by (4.2) and from (4.1) it follows that

(4.3) $$\rho_{00} + 3p_0 = \frac{\Lambda}{4\pi}, \quad \text{and} \quad R_{ij} = 4\pi\left(\rho_{00} - p_0 + \frac{\Lambda}{4\pi}\right) g_{ij}$$

which is the necessary and sufficient condition for a 3-space to possess constant Riemannian curvature.[6] It is also well known that an isotropic system of coördinates exists for the Einstein static universe.

In the general case when p_0 is not constant, a necessary and sufficient condition for (4.1) to admit isotropic solutions is still (2.7). We can substitute R_{ij} and R from (4.1) into (2.7) and take into account relations (2.9) and (2.10). Finally we obtain an equation involving the covariant derivatives of U analogous to (2.11). By a similar process which was used in deriving (2.12) we find

(4.4) $$U_{i,j} = \alpha(U_iU_j - \tfrac{1}{3}g_{ij}U^hU_h) - 2\pi\left(\rho_{00} + 3p_0 - \frac{\Lambda}{4\pi}\right)\left(g_{ij} - \frac{U_iU_j}{U^hU_h}\right) U,$$

where α plays the same rôle as α in (2.12).

The condition of integrability (2.9) for $U_{i,j}$ in (4.4) must be satisfied. We thus arrive at the conclusion that the functions α, ρ_{00}, U^hU_h must be all functions of U alone and the derivatives of α and ρ_{00} are related by

(4.5) $$\tfrac{1}{3}\left(\frac{d\alpha}{dU} - \frac{\alpha}{U} + \frac{\alpha^2}{3}\right) U^hU_h - 2\pi(\rho_{00} + p_0) - \frac{8\pi}{3}\left(\rho_{00} + 3p_0 - \frac{\Lambda}{4\pi}\right) U\alpha$$
$$- 4\pi^2\left(\rho_{00} + 3p_0 - \frac{\Lambda}{4\pi}\right)^2 \frac{U^2}{U^hU_h} + 2\pi U\frac{d\rho_{00}}{dU} = 0.$$

If we contract (4.4) by U^i, we obtain

(4.6) $$\frac{d}{dU}(U^hU_h) = \tfrac{4}{3}\alpha U^hU_h$$

which has the solution,

(4.7) $$U^hU_h = -k^2 \exp(\tfrac{4}{3}\textstyle\int \alpha dU),$$

where k^2 is a constant of integration. If ρ_{00} and p_0 are related in the form of an equation of state as in the theory of gases, then equations (4.2), (4.5) and (4.7) can be solved completely for the functions ρ_{00}, p_0, α and U^hU_h in terms of U. The above results can be summarized in the following theorem:

[6] J. A. Schouten and D. J. Struik, *American Journal of Mathematics*, vol. 43 (1921), p. 214; or inserting R_{ij} from (4.3) in (2.10) we obtain the same result (R. G. 84).

THEOREM. *If in* (4.1) ρ_{00} *is a given function of* p_0, *then a necessary and sufficient condition for* (4.1) *to admit an isotropic solution is* (4.4) *and* $R = 16\pi(\rho_{00} + \Lambda/4\pi)$ *where* α, $U^h U_h$ *and* p_0 *are determined by* (4.2), (4.5) *and* (4.7).

We have sketched the proof for the necessary condition of the theorem. The sufficient condition can be established as in the case for the field in empty space.

The general results of the paper were obtained when the author was in Peiping before the Summer of 1936 and the details have been completed here in Princeton.

It is a pleasure for him to express his gratitude to National Tsing Hua University for granting the sabbatical leave 1936-1937 and to the Institute for Advanced Study and Princeton University for their hospitality and the facilities of research put at his disposal.

INSTITUTE FOR ADVANCED STUDY,
PRINCETON, N. J.

ON TENSORS RELATIVE TO THE EXTENDED POINT TRANSFORMATION.*

By H. V. Craig.

1. Introduction. The transformation equations of the quantities dx^r/dt and $\partial f/\partial x^r$ may be said to lead to the tensor concept. Quite similarly the generalizations of these vectors, namely, $x^{(a+1)r}$, $\partial f(x \cdot x', \cdot \cdot \cdot, x^{(M)})/\partial x^{(a)r}$, $(x^{(a)r} = d^a x^r/dt^a)$ suggest a development which at first sight appears to be an extension of this very important notion. More properly perhaps the new quantities "extensors" should be regarded merely as special tensors—tensors relative to a rather special transformation. However, because of their special character a process occurs (contraction over a reduced range) which does not appear in ordinary tensor analysis.

No problem is involved in incorporating the developments of ordinary tensor analysis and consequently our interest lies elsewhere. Two subjects which are perhaps worthy of investigation are: (a) the systemization of certain of the results appearing in the various higher order geometries,[1] for example, we shall show that most of the invariants given may be regarded as formed by contraction over a reduced range; (b) the construction of tensors from extensors.

2. Notation. Following Schouten we shall use but one root letter x for all coördinate systems and distinguish between different coördinate systems by means of the letters employed as indices. We shall have occasion to consider simultaneously at most three different coördinate systems, let us call them temporarily x, y, and z, and shall adopt the associations: $x : r, s, t$; $y : i, j, k$; $z : u, v, w$. Thus, $x^r = x^r$; $x^i = y^i$; $x^u = z^u$. Likewise if V is a vector then V^r will stand for the r-th component of V in the x coördinate system, while V^i will denote the i-th component of V in the y system etc. Similarly, with regard to tensors and extensors, the presence as an index of one or more of

* Received March 10, 1937; Revised (notation only) July 13, 1937.

[1] I. e. geometries whose metrics depend on derivatives of higher order than the first. A bibliography of papers in this field is appended. The geometry studied by J. L. Synge is apparently the most general being based on a metric $F(t, x, x', \cdot \cdot \cdot, x^{(M)})$. A. Kawaguchi and others have investigated metrics which do not depend on t explicitly, whereas, the present writer has confined his attention primarily to the very special case in which Fdt is invariant under parameter change.

the letters i, j, k will serve to indicate that the component involved belongs to the y coördinate system etc.

Derivatives with respect to the parameter t will be designated for the most part by means of enclosed indices, whereas partial derivatives will be indicated by subscripts, specifically:

$$
\begin{aligned}
&x'^r = dx^r/dt; \quad x^{(a)r} = d^a x^r/dt^a; \quad F_{(a)r} = \partial F/\partial x^{(a)r}; \\
(2.1) \qquad &X_i{}^r = X^{(0)r}_{(0)i} = \partial x^r/\partial x^i; \quad X^{(a)r}_{(\gamma)i} = \partial x^{(a)r}/\partial x^{(\gamma)i}; \\
&X_i{}^{r(a)} = d^a X_i{}^r/dt^a;
\end{aligned}
$$

$\binom{u}{v}$ is a binomial coefficient.

In addition to the dimensionality of the space, N, we shall require another invariant M—the order of the space. The summations indicated by repeated lower case Latin indices shall be from one to N, whereas those represented by a repetition of small Greek indices shall be from zero to M unless otherwise specified. Whenever the range of summation of a Greek index is not zero to M a summation sign will be introduced and the range indicated in the usual way. It should be borne in mind that quantities such as $X^{(a)r}_{(\beta)i}$ have the value zero if $\alpha < \beta$ and consequently there is some latitude in the range of summations involving these derivatives. Incidentally, in all but a very few instances the irregularities of range could be removed by defining all symbols bearing negative indices to have the value zero.

Following McConnell we shall use *capital* letters for repeated indices which are *not* to be summed.

Finally, the various scalars, vectors, and extensors which are to appear will be assumed to be absolute rather than relative and to be functions of the coördinate variables and their derivatives with respect to the parameter t.

3. Extensors. The point transformation: $x^r = x^r(y^1, y^2, \cdots, y^N)$, which we assume to be of class ω and regular,[2] gives rise upon successive differentiation to the "extended point transformation":

$$
(3.1) \qquad \begin{cases} x^r = x^r(y); \quad x'^r = X_i{}^r x'^i; \\ x''^r = X_i{}^r x''^i + X_{ij}{}^r x'^i x'^j; \quad x^{(M)r} = X_i{}^r x^{(M)i} + \cdots; \end{cases}
$$

and the notion of extensor is merely that of tensor relative to this extended transformation. As particular examples of these quantities and as an illustration of the notation adopted let us consider the transformation equations of $x^{(a+1)r}$ or $x'^{(a)r}$, and $F_{(a)r}$.

[2] See Veblen and Whitehead, *The Foundations of Differential Geometry*, London, Cambridge, chap. 3.

Examining (3.1) we note that $x^{(\alpha)r}$ may be expressed in terms of $x^{(\beta)i}, \beta = 0, \cdots, \alpha$, hence

$$(3.2) \qquad (x^{(\alpha)r})' = X^{(\alpha)r}_{(\beta)i}(x^{(\beta)i})'$$

or otherwise expressed

$$(3.3) \qquad x'^{r(\alpha)} = x'^{i(\beta)}X^{(\alpha)r}_{(\beta)i} = \sum_{i=1}^{N}\sum_{\beta=0}^{M} x'^{i(\beta)}X^{(\alpha)r}_{(\beta)i}; \quad (X^{(\alpha)r}_{(\beta)i} = 0, \text{ for } \beta > \alpha).$$

This result may be generalized by means of the very useful formula [3]

$$(3.4) \qquad X^{(A)r}_{(B)i} = \binom{A}{B} X_i{}^{r(A-B)}$$

thus if V is a sufficiently differentiable vector

$$(3.5) \qquad V^{r(A)} = (X_i{}^r V^i)^{(A)} = \sum_{\beta=0}^{A}\binom{A}{\beta} X_i{}^{r(A-\beta)} V^{i(\beta)} = X^{(A)r}_{(\beta)i} V^{i(\beta)}.$$

Finally, the transformation equation for $F_{(\alpha)r}$, F being a differentiable scalar, is

$$(3.6) \qquad F_{(\beta)i} = F_{(\alpha)r} X^{(\alpha)r}_{(\beta)i},$$

and we are now ready to turn to the definition of extensor confining our statement for the sake of brevity to a third order quantity.

Definition. If there is given at a point of a curve of class C^M a set of $n^3(M+1)^2$ labelled numbers or components in each coördinate system of our group and if further the components $T^{\alpha r}_{s.\beta t}$ associated with any coördinate system x^r are related to the components $T^{\gamma i}_{j.\delta k}$ belonging to any other system x^i according to the transformation equation

$$(3.7) \qquad T^{\alpha r}_{s.\beta t} = T^{\gamma i}_{j.\delta k} X^{(\alpha)r}_{(\gamma)i} X_s{}^j X^{(\delta)k}_{(\beta)t}$$

then we shall speak of these labelled numbers as the components of a mixed third order extensor—excontravariant of order one, covariant of order one, and excovariant of order one.

In particular if $V_{\alpha r} = F_{(\alpha)r}$, we shall speak of $V_{\alpha r}$ as the exgradient of F; likewise the quantities $x'^{r(\alpha)}$ might be regarded as the components of a tangent vector in the extended sense.

The questions of consistency which arise in connection with (3.7) may

[3] See H. V. Craig, 5, p. 457. Capital indices it will be recalled are not summed.

be disposed of as in ordinary tensor analysis. Obviously, the extended transformation (3.1) admits of a unique inverse and further examination reveals that the Jacobian is $(\partial x/\partial y)^{M+1}$, hence this transformation is regular. Furthermore, we may write

$$X^{(\alpha)r}_{(\beta)i}X^{(\beta)i}_{(\gamma)s}=\delta_\gamma{}^\alpha\delta_s{}^r \tag{3.8}$$

and

$$X^{(\alpha)r}_{(\gamma)u}=X^{(\alpha)r}_{(\beta)i}X^{(\beta)i}_{(\gamma)u}. \tag{3.9}$$

These formulas which are evident from the $(M+1)N$ dimensional viewpoint, can also be obtained by applying[4] (3.4) to the identity

$$X_u{}^{r(A-\Gamma)}=(X_i{}^rX_u{}^i)^{(A-\Gamma)}=\sum_{\beta=0}^{A-\Gamma}\binom{A-\Gamma}{\beta}X_i{}^{r(A-\Gamma-\beta)}X_u{}^{i(\beta)}. \tag{3.10}$$

Obviously, the rules relating to the sum and product of tensors carry over. Likewise, the contraction of an extensor yields one of lower order thus

$$T^{\alpha r}_{s.\alpha r}=T^{\gamma i}_{j.\delta k}X^{(\alpha)r}_{(\gamma)i}X_s{}^jX^{(\delta)k}_{(\alpha)r}=T^{\gamma i}_{j.\gamma i}X_s{}^j. \tag{3.11}$$

A very simple example of contraction is furnished by the formula

$$F'=F_{(\alpha)r}x'^{(\alpha)r}. \tag{3.12}$$

More, generally, if F is any differentiable scalar and V any differentiable vector then $F_{(\alpha)r}$ and $V^{r(\alpha)}$ are extensors and consequently $F_{(\alpha)r}V^{r(\alpha)}$ is an invariant. However, the quantities $F_{(\alpha)i(\beta)j}$ are not the components of an extensor, since they follow the transformation rule

$$F_{(\alpha)i(\beta)j}=F_{(\gamma)r(\epsilon)s}X^{(\gamma)r}_{(\alpha)i}X^{(\epsilon)s}_{(\beta)j}+F_{(\gamma)r}X^{(\gamma)r}_{(\alpha)i(\beta)j}. \tag{3.13}$$

Incidentally, $X^{(\Gamma)r}_{(A)i(\beta)j}$ may be written in the form $\binom{\Gamma}{A}X_i^{r(\Gamma-A)}{}_{(\beta)j}$ and consequently it vanishes whenever β exceeds $\Gamma-A$.

The matter of eliminating the second derivatives from (3.13) may be treated just as in tensor analysis. Thus, let $g_{\alpha i\beta j}$ be a second order extensor of non-vanishing determinant. The first N terms in the first row of this determinant would be the quantities $g_{01.0j}$, while the second group of N would consist of $g_{01.1j}$ etc. By forming the derivatives $g_{\alpha i.\beta j.(\gamma)k}$, Christoffel symbols and an affine connection could be constructed in the usual way. Hence, out of a scalar or extensor, extensors of any order may be constructed by covariant differentiation and finally invariants by contraction.

---

[4] First replace β with $\beta-\Gamma$ and note that the new limits of summation may be taken to be $\beta=0$, $\beta=M$.

4. Extensors of reduced range. Examining equation (3.7) and recalling that $X^{(\delta)k}_{(\beta)t}$ is equal to zero whenever β exceeds δ, we observe that the components of $T^{\alpha r}_{s.\beta t}$ which have β equal to or greater than b depend only on the similarly labelled components $T^{\gamma i}_{j.\delta k}$ i.e. on those for which δ is not less than b. The collection of all such components will be referred to as the components of an extensor of reduced range. For example, the covariant tensor $F_{(M)r}$ is an extensor of reduced range ($b = M$) and likewise the quantities $F_{(M-1)r}$, $F_{(M)r}$ ($b = M - 1$). Similarly, the range of a superscript may be reduced. Thus, returning to (3.7) we note that the $T^{\alpha r}_{s.\beta t}$ having α equal to or less than a depend only on the corresponding components $T^{\gamma i}_{j.\delta k}$, $\gamma \leqq a$. A particular example is furnished by the derivatives of a contravariant vector V, thus

$$V^{i(\beta-E)} = \sum_{\gamma=E}^{M} V^{r(\gamma-E)} X^{(\beta-E)i}_{(\gamma-E)r}; \qquad M \geqq \beta \geqq E. \tag{4.1}$$

If E is equal to M, the range is completely reduced and (4.1) becomes a tensor transformation.

Although, differentiation of equation (3.7) yields second order derivatives we can employ this process over a restricted range to construct reduced extensors. As an illustration let us consider the following theorem which obviously admits of generalization.

THEOREM (4.1). *If $T^{\alpha i}_{\beta j}$ are the components of an extensor of reduced range ($\alpha \leqq a$, $\beta \geqq b$) then for $\gamma > a$ and $\gamma > M - b$ simultaneously, the set of quantities $T^{\alpha i}_{\beta j.(\gamma)k}$ constitute the components of an extensor of reduced range.*

Proof. The law of transformation of $T^{\alpha i}_{\beta j}$ is expressed by the relationship;

$$T^{\alpha i}_{\beta j} = T^{\sigma r}_{\delta s} X^{(\alpha)i}_{(\sigma)r} X^{(\delta)s}_{(\beta)j}, \qquad \alpha \leqq a,\ \beta \geqq b. \tag{4.2}$$

Differentiating (4.2) with respect to $x^{(\gamma)k}$, $\gamma > a$, $\gamma > M - b$, we obtain the desired equality

$$T^{\alpha i}_{\beta j.(\gamma)k} = T^{\sigma r}_{\delta s.(\epsilon)t} X^{(\epsilon)t}_{(\gamma)k} X^{(\alpha)i}_{(\sigma)r} X^{(\delta)s}_{(\beta)j}, \quad \alpha \leqq a,\ \beta \geqq b,\ \gamma > a,\ \gamma > M - b, \tag{4.3}$$

at once, since $X^{(\alpha)i}_{(\sigma)r.(\epsilon)t} X^{(\epsilon)t}_{(\gamma)k}$ and $X^{(\delta)s}_{(\beta)j.(\gamma)k}$ vanish by virtue of the restrictions on α, β, and γ.

An illustration of this process is furnished by the differentiation of a tensor. For example, the quantities $T^{i}{}_{jk.(\beta)l}$, $\beta > 0$, are the components of a reduced extensor.

Finally, with regard to the differentiation of (4.2) it should be noted that if γ were permitted to take on the larger of the two values a and $M - b$ or

their common value if $a = M - b$ then there would appear in the right member of (4.3) one or more terms involving mixed partial derivatives of the unprimed variables. These derivatives could be eliminated by means of a zero-th order affine connection—the resulting quantities are however irregular and of no apparent interest.

5. Contraction over a reduced range. The existence of extensors of reduced range suggests the possibility of constructing invariants and tensors by contraction over a reduced range. As a first trial it would be natural to consider the expression $\sum_{\beta=E}^{M} F_{\beta i} V^{\beta-E,i}$. Investigation however reveals that a factor $\binom{\beta}{E}$ is needed. The general process is exemplified by the theorem which follows.

THEOREM (5.1). *If $T^{ai.\beta-E,j}_{k.\ \ \gamma l}$ is an extensor of reduced range: $E \leq \beta \leq M$, $E \leq \gamma \leq M$, then the quantities $\sum_{\beta=E}^{M} \binom{\beta}{E} T^{ai.\beta-E,j}_{k.\ \ \beta j}$ are the components of a mixed extensor—covariant of order one and excontravariant of order one.*

Proof. Evidently the transformation equation to be investigated is

$$\sum_{\beta=E}^{M} \binom{\beta}{E} T^{ai.\beta-E,j}_{k.\ \ \beta j} = \sum_{\beta=E}^{M} \sum_{\delta=E}^{M} \sum_{\rho=E}^{M} T^{\gamma r.\delta-E,s}_{t.\ \ \rho q} X^{(a)i}_{(\gamma)r} X_k{}^t \binom{\beta}{E} X^{(\beta-E)j}_{(\delta-E)s} X^{(\rho)q}_{(\beta)j}. \tag{5.1}$$

Now by virtue of (3.4) and the equality of $\binom{B}{E}\binom{B-E}{\Delta-E}\binom{B}{\Delta}^{-1}$ and $\binom{\Delta}{E}$, we may write successively,

$$\binom{B}{E} X^{(B-E)j}_{(\Delta-E)s} = \binom{B}{E}\binom{B-E}{\Delta-E} X_s{}^{j(\beta-\Delta)} = \binom{B}{E}\binom{B-E}{\Delta-E}\binom{B}{\Delta}^{-1} X^{(B)j}_{(\Delta)s} = \binom{\Delta}{E} X^{(B)j}_{(\Delta)s} \tag{5.2}$$

and consequently

$$\sum_{\beta=E}^{M} \binom{\beta}{E} X^{(\beta-E)j}_{(\Delta-E)s} X^{(\rho)q}_{(\beta)j} = \sum_{\beta=E}^{M} \binom{\Delta}{E} X^{(\beta)j}_{(\Delta)s} X^{(\rho)q}_{(\beta)j}. \tag{5.3}$$

For $\Delta \geqq E$ the range of β may be taken to be 0 to M and hence the last summation sign may be dropped. Thus the last member of (5.3) reduces to $\binom{\Delta}{E}\delta_\Delta{}^\rho \delta_s{}^q$ and the right member of (5.1) becomes

$$\sum_{\delta=E}^{M} \binom{\delta}{E} T^{\gamma r.\delta-E,s}_{t.\ \ \delta s} X^{(a)i}_{(\gamma)r} X_k{}^t \tag{5.4}$$

and the theorem is established.

The quotient law for contraction over a reduced range is illustrated by the next theorem.

THEOREM (5.2). *If for each differentiable vector V the set of quantities $\sum_{\beta=E}^{M}\binom{\beta}{E}T^{j}{}_{ai.\beta k}V^{k(\beta-E)}$ are the components of a mixed extensor of the type indicated by the free indices then $T^{j}{}_{ai.\beta k}$ is a third order mixed extensor of the reduced range $E\leq\beta\leq M$.*

Proof. By virtue of the hypotheses we have

$$(5.5)\qquad \begin{aligned}&\sum_{\beta=E}^{M}\binom{\beta}{E}T^{j}{}_{ai.\beta k}V^{k(\beta-E)}\\ &=\sum_{\delta=E}^{M}\binom{\delta}{E}T^{s}{}_{\gamma r.\delta t}V^{t(\delta-E)}X_{s}{}^{j}X^{(\gamma)r}_{(a)i}\\ &=\sum_{\beta=E}^{M}\sum_{\delta=E}^{M}\binom{\delta}{E}T^{s}{}_{\gamma r.\delta t}V^{k(\beta-E)}X^{(\delta-E)t}_{(\beta-E)k}X_{s}{}^{j}X^{(\gamma)r}_{(a)i}\end{aligned}$$

and consequently

$$(5.6)\qquad \sum_{\beta=E}^{M}\left[\binom{\beta}{E}T^{j}{}_{ai.\beta k}-\sum_{\delta=E}^{M}\binom{\delta}{E}T^{s}{}_{\gamma r.\delta t}X^{(\delta-E)t}_{(\beta-E)k}X_{s}{}^{j}X^{(\gamma)r}_{(a)i}\right]V^{k(\beta-E)}\equiv 0.$$

By substituting for V^{k} successively the quantities $\delta_{l}{}^{k}$, $\delta_{l}{}^{k}t$ etc., t being the parameter, we learn that the bracket in (5.6) must vanish for each admissible value of k and β. Furthermore, by way of (5.2) we may replace $\binom{\Delta}{E}X^{(\Delta-E)t}_{(B-E)k}$ by $\binom{B}{E}X^{(\Delta)t}_{(B)k}$ and the theorem follows.

6. The structure of certain tensors. The most important scalar equations of generalized geometry are due to Zermelo and express the various conditions for the invariance of Fdt, $F=F(x,x',\cdots,x^{(M)})$, under change of parameter. It may be of interest to examine their structure in the light of the foregoing developments. The equations[5] are:

$$(6.1)\qquad (x'^{r}T)^{(a)}F_{(a)r}=(FT)';$$

$$(6.2)\qquad \sum_{a=\Gamma}^{M}\binom{a}{\Gamma}x'^{r(a-\Gamma)}F_{(a)r}=\delta_{\Gamma}{}^{1}F;$$

$$(6.3)\qquad \sum_{a=1}^{M}\binom{a}{1}x'^{r(a-1)}Z_{ar}=F;$$

$$(6.4)\qquad \left(Z_{Ar}=1/A\cdot\sum_{\beta=A}^{M}(-1)^{\beta-A}F_{(\beta)r}{}^{(\beta-A)},\qquad (A=1,\cdots,M)\right).$$

[5] See A. Kneser, *Lehrbuch der variationsrechnung* (1900), p. 195. Also, H. V. Craig, 1, p. 559; 5, p. 461; 6, p. 835.

The left members of the first two of the foregoing set are obviously examples of complete contraction and contraction over a reduced range. The third equation suggests the following theorem.

THEOREM (6.1). *The set of quantities $Z_{\alpha r}$, $1 \leq \alpha \leq M$, constitute the components of an extensor of reduced range.*

Proof. One verifies readily that for the ranges: $\alpha = M$; $M - 1 \leq \alpha \leq M$; the quantities $Z_{\alpha r}$ are indeed the components of a reduced extensor and we proceed to establish the theorem by induction. Thus, we shall assume the theorem for the range $A + 1 \leq \alpha \leq M$ and on this basis deduce that Z_{Ar} obeys the desired transformation equation. Now it may be shown that these quantities obey the recursion formula [6]

$$Z_{Ai} = 1/A \cdot F_{(A)i} - (A+1)/A \cdot Z_{A+1,i}' \tag{6.5}$$

and consequently it will suffice to investigate the transformation of the right member of (6.5).

By virtue of the extensor character of $F_{(A)i}$ and $Z_{A+1,i}$ we may write

$$F_{(A)i} = \sum_{\gamma=A}^{M} F_{(\gamma)r} X^{(\gamma)r}_{(A)i}, \tag{6.6}$$

$$Z_{A+1,i} = \sum_{\gamma=A+1}^{M} Z_{\gamma r} X^{(\gamma)r}_{(A+1)i} \tag{6.7}$$

and, differentiating (6.7) with respect to the parameter t,

$$Z_{A+1,i}' = \sum_{\gamma=A+1}^{M} Z_{\gamma r}' X^{(\gamma)r}_{(A+1)i} + \sum_{\gamma=A+1}^{M} Z_{\gamma r} X^{(\gamma)r}_{(A+1)i}{}'. \tag{6.8}$$

If we replace γ with $\gamma + 1$ in the first term of the right member of (6.8) we get

$$Z_{A+1,i}' = \sum_{\gamma=A}^{M-1} Z_{\gamma+1,r}' X^{(\gamma+1)r}_{(A+1)i} + \sum_{\gamma=A+1}^{M} Z_{\gamma r} X^{(\gamma)r}_{(A+1)i}{}'. \tag{6.9}$$

Furthermore, by way of (6.5), (6.6), and (6.9) we have

$$\begin{aligned} Z_{Ai} = 1/A \sum_{\gamma=A}^{M} F_{(\gamma)r} X^{(\gamma)r}_{(A)i} &- (A+1)/A \sum_{\gamma=A}^{M-1} Z_{\gamma+1,r}' X^{(\gamma+1)r}_{(A+1)i} \\ &- (A+1)/A \sum_{\gamma=A+1}^{M} Z_{\gamma r} X^{(\gamma)r}_{(A+1)i}{}'. \end{aligned} \tag{6.10}$$

Evidently, the lower limit of the last summation can be changed to A while

[6] See H. V. Craig, 1, p. 560.

$$-(A+1)/A \cdot X^{(\gamma+1)r}_{(A+1)i} \quad \text{can be replaced by} \quad -(\gamma+1)/A \cdot X^{(\gamma)r}_{(A)i},$$

$-(A+1)/A \cdot X^{(\gamma)r}_{(A+1)i}{}'$ by $(1-\gamma/A)X^{(\gamma)r}_{(A)i}$, and Z_{Mr} by $1/M \cdot F_{(M)r}$. This done the right member of (6.10) assumes the form

$$(6.11)\quad \sum_{\gamma=A}^{M-1}\{\gamma/A[1/\gamma \cdot F_{(\gamma)r} - (\gamma+1)/\gamma \cdot Z_{\gamma+1,r}{}'] + (1-\gamma/A)Z_{\gamma r}\}X^{(\gamma)r}_{(A)i} + (1/A \cdot F_{(M)r} + (1-M/A)1/M \cdot F_{(M)r})X^{(M)r}_{(A)i}$$

and thus by way of (6.5) equation (6.10) reduces to

$$(6.12)\qquad Z_{Ai} = \sum_{\gamma=A}^{M-1} Z_{\gamma r}X^{(\gamma)r}_{(A)i} + Z_{Mr}X^{(M)r}_{(A)i}$$

and our theorem is proved and the structure of (6.3) established.

As a final scalar let us consider the Synge invariant;[7]

$$(6.13)\qquad \sum_{\alpha=0}^{M-1} V^{r(\alpha)} \sum_{\beta=1}^{M-\alpha} (-1)^{\beta-1} F_{\cdot\cdot(\alpha+\beta)r}^{(\beta-1)}.$$

If we replace first α with $\alpha-1$ and then β by $\beta-\alpha+1$ the expression (6.13) may be written in the form

$$(6.14)\qquad \sum_{\alpha=1}^{M} \binom{\alpha}{1} V^{r(\alpha-1)} \cdot 1/\alpha \cdot \sum_{\beta=\alpha}^{M} (-1)^{\beta-\alpha} F_{\cdot\cdot(\beta)r}^{(\beta-\alpha)}$$

which by virtue of the definition of $Z_{\alpha r}$ reduces to the contraction

$$(6.15)\qquad \sum_{\alpha=1}^{M} \binom{\alpha}{1} V^{r(\alpha-1)} Z_{\alpha r}.$$

Among the tensors whose character is not apparent from the N-dimensional viewpoint are the quantities $\overset{p-R}{D}_{rs}(T)V^s$, due to Kawaguchi,[8]

$$(6.16)\qquad \overset{p-R}{D}_{rs}(T)V^s = \sum_{\alpha=R}^{p} \binom{\alpha}{R} T_{r(\alpha)s} V^{s(\alpha-R)}.$$

This is obviously another example of contraction over a reduced range.

Among the most interesting quantities of higher order geometry are the vectors $\overset{A}{E}_r$ [9]

$$(6.17)\qquad \overset{A}{E}_r = \sum_{\gamma=a}^{M} (-1)^{\gamma} \binom{\gamma}{A} F_{\cdot\cdot(\gamma)r}^{(\gamma-A)}.$$

Examination of their structure suggests that these vectors may be generalized

---

[7] See J. L. Synge, 1, p. 683.

[8] See A. Kawaguchi, 4, p. 149.

[9] See J. L. Synge, 1, p. 684; H. V. Craig, 6, p. 833.

by replacing the exgradient $F_{(\gamma)r}$ with an arbitrary differentiable extensor $V_{\gamma r}$. As an aid to this investigation we introduce the identity

$$(6.18) \qquad \sum_{\mu=A}^{\Gamma} \binom{\mu}{A} (-1)^{\mu} (F X^{(\Gamma)r}_{(\mu)i})^{(\mu-A)} = \binom{\Gamma}{A} (-1)^{\Gamma} F^{(\Gamma-A)} X_i^r$$

which may be established as follows.

By virtue of the rule for differentiating a product and the formula (3.4) we may express the left member of (6.18) in the form

$$(6.19) \qquad \sum_{\mu=A}^{\Gamma} \binom{\mu}{A} (-1)^{\mu} \sum_{\sigma=0}^{\mu-A} \binom{\mu-A}{\sigma} F^{(\sigma)} \binom{\Gamma}{\mu} X_i^{r(\Gamma-A-\sigma)}.$$

Interchanging the summation signs and replacing $\binom{M}{A}\binom{M-A}{\Sigma}\binom{\Gamma}{M}$ by the equivalent product $\binom{\Gamma}{A}\binom{\Gamma-A}{\Sigma}\binom{\Gamma-A-\Sigma}{\Gamma-M}$, we obtain the expression,

$$(6.20) \qquad \sum_{\sigma=0}^{\Gamma-A} \binom{\Gamma}{A}\binom{\Gamma-A}{\sigma} F^{(\sigma)} X_i^{r(\Gamma-A-\sigma)} \sum_{\mu=\sigma+A}^{\Gamma} (-1)^{\mu} \binom{\Gamma-A-\sigma}{\Gamma-\mu}.$$

Now, if we replace μ with $\mu+\sigma+A$ and denote $\Gamma-A-\sigma$ by b, the last summation may be written in the form $(-1)^{\sigma+A} \sum_{\mu=0}^{b} (-1)^{\mu} \binom{b}{\mu}$ and consequently reduces to $(-1)^{\sigma+A}\delta_0^b$. Hence (6.20) reduces to $(-1)^{\Gamma}\binom{\Gamma}{A} F^{(\Gamma-A)} X_i^r$, the right member of (6.18). This accomplished we are ready to consider the theorem:

THEOREM (6.2). *If $V_{\gamma r}$ is a differentiable extensor then the quantities*

$$(6.21) \qquad \sum_{\gamma=A}^{M} (-1)^{\gamma} \binom{\gamma}{A} V^{(\gamma-A)}_{\gamma r}$$

are the components of a vector.

Proof. If in (6.18) we replace Γ with γ, F with $V_{\gamma r}$ and sum both members from $\gamma = A$ to $\gamma = M$, we obtain (after changing the order of summation on the left)

$$(6.22) \qquad \sum_{\mu=A}^{M} (-1)^{\mu} \binom{\mu}{A} \left(\sum_{\gamma=\mu}^{M} V_{\gamma r} X^{(\gamma)r}_{(\mu)i} \right)^{(\mu-A)} = \sum_{\gamma=A}^{M} (-1)^{\gamma} \binom{\gamma}{A} V^{(\gamma-A)}_{\gamma r} X_i^r$$

or

$$(6.23) \qquad \sum_{\mu-A}^{M} (-1)^{\mu} \binom{\mu}{A} V^{(\mu-A)}_{\mu i} = \sum_{\gamma=A}^{M} (-1)^{\gamma} \binom{\gamma}{A} V^{(\gamma-A)}_{\gamma r} X_i^r,$$

and our theorem is established.

TEXAS UNIVERSITY.

BIBLIOGRAPHY.

E. Cartan, "La géométrie de l'integral $\int F(x, y, y', y'')\,dx$," *Journal de mathématiques pures et appliques*, vol. 15 (1936), p. 42.

H. V. Craig, 1, "On the solution of the Euler equations for their highest derivatives," *Bulletin of the American Mathematical Society*, vol. 36 (1930), p. 560; 2, "On parallel displacement in a non-Finsler space," *Transactions of the American Mathematical Society*, vol. 33 (1931), p. 129; 3, "On a covariant differentiation process," *Bulletin of the American Mathematical Society*, vol. 37 (1931), p. 731; 4, "On a covariant differentiation process, II," *Bulletin of the American Mathematical Society*, vol. 39 (1933), p. 919; 5, "On a generalized tangent vector," *The American Journal of Mathematics*, vol. 57 (1935), p. 457; 6, "On a generalized tangent vector, II," *The American Journal of Mathematics*, vol. 58 (1936), p. 833.

S. Hokari, "Die Geometrie des Integrals $\int (a_i a_j x''^i x''^j + 2ba_i x''^i + c)^{1/p} dt$," *Proceedings of the Imperial Academy of Tokio*, vol. 12 (1936), p. 209.

H. Hombu, 1, "On a non-Finsler metric space," *The Tôhoku Mathematical Journal*, vol. 37 (1933), p. 190; 2, "Invariantentheorie des Integrals $\int F(x, y, y', y'', y''')\,dx$," *Proceedings of the Imperial Academy of Tokio*, vol. 12 (1936), p. 156.

A. Kawaguchi, 1, "Theory of connections in the generalized Finsler Manifold," *Proceedings of the Imperial Academy of Tokio*, vol. 7 (1931), p. 211; 2, "Die Differentialgeometrie in der varallgemeinerten Mannigfaltigkeit," *Rendiconti Circolo Matematico di Palermo*, vol. 56 (1932), p. 245; 3, "Theory of connections in the generalized Finsler Manifold, III," *Proceedings of the Imperial Academy of Tokio*, vol. 9 (1933), p. 347; 4, "Some intrinsic derivations in a generalized space," *Proceedings of the Imperial Academy of Tokio*, vol. 12 (1936), p. 149; 5, "Certain identities in a generalized space," *Proceedings of the Imperial Academy of Tokio*, vol. 12 (1936), p. 152; 6, "Die Geometrie des Integrals $\int (A_i x''^i + B)^{1/p} dt$," *Proceedings of the Imperial Academy of Tokio*, vol. 12 (1936), p. 205.

D. D. Kosambi, "Path-spaces of higher order," *The Quarterly Journal of Mathematics*, Oxford Series, vol. 7 (1936), p. 97.

J. L. Synge, "Some intrinsic and derived vectors in a Kawaguchi space," *The American Journal of Mathematics*, vol. 57 (1935), p. 679.

A SERIES OF INVOLUTORIAL CREMONA TRANSFORMATIONS IN S_n BELONGING MULTIPLY TO A NON-LINEAR LINE COMPLEX.*

By Virgil Snyder and Evelyn Carroll-Rusk.

The involutorial transformation I will first be developed in three way projective space S_3, then extended to S_n.

1. Definition of the involution. Consider a pencil

$$\mu_2 H_1(x) - \mu_1 H_2(x) = 0 \tag{1}$$

of quadric surfaces, with a proper or composite curve C_4 for base, and a rational ruled surface R_{n+m} of order $n+m$, generated by the line

$$(ax) = 0, \qquad (bx) = 0, \tag{2}$$

in which each coefficient a_i is a binary form in λ_1, λ_2 of order m_1, and each coefficient b_i is a binary form in λ_1, λ_2 of order m_2.

Let ϕ_1, ϕ_2 be two binary forms in μ_1, μ_2 of degree k, and ψ_1, ψ_2 be two quadratic forms in λ_1, λ_2.

Between the ϕ, ψ is the relation

$$\phi_1/\phi_2 = \psi_1/\psi_2. \tag{3}$$

With these premises the involution I can be defined as follows: Given any point (y). By (1) it determines a quadric of the pencil, hence $\mu_1 = H_1(y)$, $\mu_2 = H_2(y)$.

These values substituted in (3) define λ_1/λ_2 as roots of a quadratic equation and consequently, by (2), a pair of generators of R_{n+m} belonging to a $g_2{}^1$.

Through y can be drawn one transversal meeting both, and it will meet the quadric $H(y)$ determined by (y) in a second point (y'), conjugate of (y) in I.

2. Properties of the involution. Fundamental elements. From (3) the quadrics of the pencil (1) are arranged in sets of k, all associated with the same pair of generators of $R_{m_1+m_2}$. Every quadric of the pencil is invariant

* Presented to the American Mathematical Society at New York, March 27, 1937. The first part was also read before the International Congress of Mathematicians, Oslo, July 13, 1936. Received by the Editors May 3, 1937.

under the involution I, and every line joining a pair PP' of conjugate points contains k such pairs. For each set of k quadrics the transformation is expressed by means of a linear line congruence.

The fundamental elements are of three kinds:

(a) the basic quartic C_4 of the pencil of quadrics.

From a point P on C_4 can be drawn a transversal t to every pair of conjugate generators of $R_{m_1+m_2}$. Each such transversal determines a set of k quadrics, and each quadric has one point on the transversal apart from P. The locus of this point as t describes a cone with vertex P is the image of P. As P describes C_4, the curve generates a surface K, the principal element, image of the curve C_4.

(b) Let g_1, g_2 be the generators of $R_{m_1+m_2}$ associated with the quadrics $H_1, \cdots, H_k$. Each line meets all k associated quadrics each in two points. Let Q be a point g_1, H_i. The plane Q, g_2 meets H_i in a conic passing through Q; this conic is the image of Q. As g describes $R_{m_1+m_2}$ Q describes a curve γ which is double on the web of surfaces conjugate to the planes of space, and it lies on the surface of invariant points. As Q describes γ, its conjugate conic describes the principal surface Γ.

These are the only fundamental curves of the first kind. The surfaces K, Γ comprise the Jacobian of the web.

The sets of k quadrics associated with coincident lines $g_1 = g_2$ contain an infinite number of parasitic conics, hence appear as factors and do not affect the proper transformation.

(c) The common bisecants of γ and C_4 are parasitic. Each point of every such line is transformed into the whole line. All these lines lie on the surface σ of invariant points.

(d) For certain values of μ the associated lines g_1, g_2 intersect. Their plane meets the associated k quadrics in k conics of a pencil, each of which is parasitic. These conics all lie on the surface σ of invariant points.

The lines (c) and the conics (d) include all the fundamental curves of the second kind.

3. Analytic procedure. Let the line t meet g_1 and z' and g_2 in z''. By (3) the coördinates of $z = z' + z''$ can be expressed rationally in terms of μ. Each z_i is of order $2p + 1$ in (y) and contains μ to multiplicity p, where $p = k(m_1 + m_2 - 1)$. Then $y'_i = \sigma z_i + \tau y_i$, wherein

$$\sigma = 2(H_1(y)H_2(y,z) - H_2(y)H_1(y,z)),$$
$$\tau = -(H_1(y)H_2(z) - H_2(y)H_1(z)),$$
$$H_i(y,z) = H_i(z,y) \text{ is the polar of } H(y) \text{ as to } z.$$

The surface $\sigma = 0$ is the locus of points invariant under I. Each point of σ is a point of contact of a line of the linear congruence g_1, g_2 which touches an associated quadric.

The equations of the curve γ can be found by eliminating λ, λ' between (2) and (3). From (1), (3) may be written in the form

$$\{A_1\phi_1(H_1, H_2) + B_1\phi_2(H_1, H_2)\}\lambda_1^2 \\ + \{A_2\phi_1(H_1, H_2) + B_2\phi_2(H_1, H_2)\}\lambda_1\lambda_2 + \cdots = 0.$$

Eliminate λ between this and $(ax) = 0$. The result is a surface $F_{2km_1+2} : C_4{}^{km_1}$. Similarly, by eliminating λ between the same quadratic equation and $(bx) = 0$, we obtain $F_{2km_2+2} : C_4{}^{km_2}$. Each of these surfaces contains γ and C_4;

$$(2km_1 + 2)(2km_2 + 2) - 4k^2m_1m_2 = 4[k(m_1 + m_2) + 1],$$

but each line of R has two points on γ, hence the order of γ is $2p + 2 + 2k$. It lies on the surface $R_{m_1+m_2}$.

Let $(ax)(by) - (ay)(bx) = \Sigma P_ix_i = 0$, and $\Sigma P'_ix_i = 0$ be the equations of t, wherein $P_i = \Sigma p_{ik}y_k$, $p_{ik} = a_ib_k - a_kb_i$. This line meets $\Sigma x_i = 0$ in the point v, each coördinate v_i being obtained from the matrix

$$\begin{Vmatrix} P_1 & P_2 & P_3 & P_4 \\ P'_1 & P'_2 & P'_3 & P'_4 \\ 1 & 1 & 1 & 1 \end{Vmatrix}$$

by suppressing the i-th column. By means of (1) each coördinate can be expressed in the form

$$\begin{vmatrix} c_0\lambda_1 + c_1\lambda_2 & d_0\lambda_1 + d_1\lambda_2 \\ c_0\lambda_1' + c_1\lambda'_2 & d_0\lambda'_1 + d_1\lambda'_2 \end{vmatrix} = \begin{vmatrix} c_0 & c_1 \\ d_0 & d_1 \end{vmatrix} \cdot \begin{vmatrix} \lambda_1 & \lambda'_1 \\ \lambda_2 & \lambda'_2 \end{vmatrix},$$

hence $\lambda_1\lambda'_2 - \lambda_2\lambda'_1$ is removed as a factor from each v_i, and hence from the transformation.

The discriminant of (1) is quadratic in $\phi_1 : \phi_2$, hence there are two sets of k quadrics which appear once for all as factors. The coincident directrices are not fundamental in the transformation. The curve C_4 meets R in $4(m_1 + m_2)$ points, each of which is an actual k fold point on γ. These count as $2^k(k-1)(m_1 + m_2)$ actual double points, which must be added to the number of apparent double points of γ to determine the number of common bisecants of γ and C_4. The number of common bisecants u is $4p + 8$.

Let a common bisecant of C_4, γ meet the latter in P, P'. Through each

of these points passes a generator g, g' of R. Then g, g' are associated directrix lines associated with the quadric of the pencil on which PP' lies. Every common bisecant is a fundamental line of the second kind.

The number of parasitic conics is $p - k$, each appearing as a double base curve on the web of surfaces conjugate to the planes of space.

Each quadric of the pencil (1) is invariant under I. It contains C_4 simply but does not contain γ nor any parasitic curve.

4. Table of characteristics.

$$\begin{aligned}
&S_1 \sim S_{4p+5} : C_4{}^{2p+1}\ \gamma^2{}_{2p+2+2k}(p-k)C_2{}^2(4p+8)u\\
&C_4 \sim K_{8p+8} : C_4{}^{4p+1}\ \gamma^4{}_{2p+2+2k}(p-k)C_2{}^4(4p+8)u^2\\
&\gamma \sim \Gamma_{8p+8} : C_4{}^{4p+2}\ \gamma^3{}_{2p+2+2k}(p-k)C_2{}^4(4p+8)u^2\\
&\sigma \sim \sigma_{2p+4} : C_4{}^{p+1}\ \gamma_{2p+2+2k}(p-k)C_2(4p+8)u.\\
&J = K\Gamma.
\end{aligned}$$

5. The complex *PP'*. The line joining a point P to its conjugate P' in I contains k such pairs P, P'. The equations are

$$(ax)(by) - (ay)(bx) = 0, \qquad (a'x)(b'y) - (a'y)(b'x) = 0,$$

which may be written in the form

$$\Sigma r_{ik} p_{ik} = 0, \qquad \Sigma r'_{ik} p_{ik} = 0, \qquad r = a_i b_k - a_k b_i, \qquad p_{ik} = x_i y_k - x_k y_i.$$

Each r_{ik} is a polynomial of degree $m_1 + m_2$ in λ_1, λ_2, r'_{ik} is the same polynomial in λ'_1, λ'_2. The equations may be written

$$c_0 \lambda^{m_1+m_2} + \cdots + c_{m_1+m_2} = 0, \qquad c_0 \lambda'^{m_1+m_2} + \cdots + c_{m_1+m_2} = 0,$$

in which each c_i is linear in p_{ik}, the Plücker coördinates of the line PP'. They may be replaced by two others, each symmetric in λ, λ'. Then by (1) these can be replaced by equations in ϕ_1/ϕ_2. The ϕ eliminant equated to zero is the equation of the complex. It is not a function of k. The order of the complex is $m_1 + m_2 - 1$. It can be linear only when $m_1 = m_2 = 1$.

Every point of R is singular on the complex. Let P be any point of R, g_1 the generator of R passing through it, and g_2 the conjugate of g_1 on R. The pencil P, g_2 belongs to the complex. Similarly, every tangent plane to R is a singular plane. Among the lines of the congruence g_1, g_2 there are two which also belong to another, say g'_1, g'_2. The common transversals of these

four lines are all double lines of the complex, each containing $2k$ pairs of conjugate points.

The congruence of double lines is self dual. Its order is equal to the number of double generators of a general complex cone. Since R and g'_2 are rational, this number is the maximum for a cone of its order.

If g_1, g_2 are fixed and g'_1, g'_2 describes R, the double lines generate a ruled surface having g_1, g_2 for $(m_1 + m_2)$-fold directrices; the whole congruence of double lines is arranged on a system of these ruled surfaces. If R has a rectilinear directrix, it belongs to every ruled surface of the system; if R has two rectilinear directrices, the entire congruence of double lines is replaced by these two directrix lines. Finally, if $m_1 = m_2 = 1$, the congruence is replaced by the other regulus of R. Since each generator of this regulus contains an infinite number of pairs of conjugate points, it remains invariant under I, hence R is transformed into itself.

6. Earlier particular cases. All cases already found in the literature are included in the present paper. The case in which the two directrices are fixed belongs to the types discussed by de Paolis.[1] It can be reduced to the monoidal type. Montesano[2] considered the case $m_1 = m_2 = k = 1$ from a different stand-point. This was generalized for any k by Snyder.[3] These two are contained in a general linear complex. Various cases in which the complex PP' is composed of secants to a rational curve are given by Davis,[4] Black,[5] Carroll,[6] Dye,[7] Montague,[8] Vicianze,[9] and Schoonmaker and Snyder.[10] A decidedly more general type is given by Dye.[11]

---

[1] *Rom. Acc. Lin. Mem.* (4), vol. 1 (1885), pp. 576-608; *Rom. Acc. Lin. Rend.* (4), vol. 1 (1885), pp. 735-792, 754-758.

[2] *Rom. Acc. L. Rend.* (4), vol. 4 (1888), pp. 207-25, 277-285.

[3] *Atti Congresso intern. di Bologna*, vol. 4 (1928), pp. 13-20.

[4] *American Journal of Mathematics*, vol. 52 (1930), pp. 53-71; vol. 53 (1931), pp. 72-80; *Tôhoku Mathematical Journal*, vol. 33 (1931), pp. 254-259.

[5] *Transactions of the American Mathematical Society*, vol. 34 (1931), pp. 795-810; *Bulletin of the American Mathematical Society*, vol. 40 (1934), pp. 417-420; vol. 42 (1936), pp. 535-540.

[6] *American Journal of Mathematics*, vol. 54 (1931), pp. 707-717; vol. 56 (1934), pp. 96-108.

[7] *Transactions of the American Mathematical Society*, vol. 32 (1930), pp. 251-261; *American Journal of Mathematics*, vol. 54 (1932), pp. 499-504 (with F. R. Sharpe).

[8] *Bulletin of the American Mathematical Society*, vol. 42 (1936), pp. 727-731.

[9] *Dissertation*, Catania University (1922).

[10] *American Journal of Mathematics*, vol. 54 (1932), pp. 299-304.

[11] *Bulletin of the American Mathematical Society*, vol. 42 (1936), pp. 535-540, and a paper presented at the Duke meeting, December, 1936.

7. Extension to S_n. If in (3) each polynomial ψ_i is of degree $n-1$ in λ_1, λ_2, while in (1) and (2) the number of homogeneous variables x_i is $n+1$, the transformation I is at once extended to S_n. Instead of two skew directrix lines there are now $n-1$ linear spaces S_{n-2} defined by (2) and (3), forming an involution of order $n-1$ among the generating S_{n-2}'s of R.

Through an arbitrary point y can be drawn just one transversal t which meets all $n-1$ associated director spaces.

The fundamental elements C_4 and γ are defined as in S_3; C_4 is the base of the pencil of quadric primals, and γ is the locus of the intersection of a director S_{n-2} with its associated set of k quadric primals. The image of each point of γ is a conic cut from the primals by the plane meeting the other director spaces.

In S_n, the line through a point P on γ, in order to be parasitic must not only be a bisecant of C_4, hence a line of the quadric primal through P, but must also meet all the other director spaces S_{n-2} in points of their quadrics of intersection with their associated primals. Since every point of γ is double on $|S|$, these lines have $n-1$ double points and two points of multiplicity $2p+1$ on C_4, and are therefore base lines of the system.

The image of C_4 is a primal K, and of γ is a primal Γ. These two primals constitute the complete Jacobian of the system.

As in S_3, those values of $\mu_1, \mu_2 = \mu$ corresponding to double roots of (3) define primals of the pencil (1) which are factors of the system, and therefore do not appear as fundamental elements in the proper transformation.

In S_3 the phenomenon of a plane field of lines meeting both directrices is possible only when the directrices intersect.

In S_4, associated with every value of μ each triad of director planes determines a plane, that through the three points in which they meet by twos. This plane meets each director plane in a line, hence every line in it meets all three director planes. The plane meets its associated quadric primals in k conics, all parasitic in I. Thus, in S_4, I has ∞^1 fundamental conics of the second kind, forming a surface Σ, which is a part of the base of $|S|$, conjugate to the primes of S_4.

In S_n there are $n-1$ spaces S_{n-2} meeting by twos in $\binom{n-1}{2}$ spaces S_{n-4}. Through any point of any one of these intersections can be drawn one line meeting the others in each of the director spaces through it. The plane of these two intersecting lines meets each director space in a line. Hence for each set of k quadrics of the pencil there are ∞^{n-4} such parasitic conics,

generating a manifold Σ of $n-2$ dimensions, not lying in a linear space of dimensions less than n. It is that defined by the Veneroni transformation.[12]

These parasitic conics include all the particular cases arising from base elements lying in simpler spaces for special values of μ.

Finally, the other base elements of $|S|$ consist of the lines that are bisecants of the quartic C_4 and multisecants of γ. They form a manifold T.

The equations of the transversal line t through y meeting the director spaces are $\sum_{i=1}^{n+1} P_i^{(j)} x_i = 0$ $(j = 1, 2, \cdots, n-1)$. The point z in which t meets the prime $\Sigma x_i = 0$ is given by the matrix

$$\left\| \begin{matrix} P_1^{(1)} & \cdots & P_{n+1}^{(1)} \\ \cdots & \cdots & \cdots \\ P_1^{(n-1)} & \cdots & P_{n+1}^{(n-1)} \\ 1 & \cdots & 1 \end{matrix} \right\| .$$

By (3), each element can be reduced to a polynomial r_{ik} of degree $n-2$ in λ, the coefficients being linear in y and functions of order $(m_1 + m_2) - (n-2)$ in ϕ_1, ϕ_2. Each coördinate z_i is expressed as a determinant of order $n-1$ with elements r_{ik}. But this is the product of two determinants, one in the coefficients, and the other is $|\lambda_1^{n-2}, \lambda_2^{n-1}, \cdots, 1| = \pi(\lambda_i - \lambda_k)$.

This second factor is the same for all z_i, and divides out of the transformation. This explains why the coincident director spaces do not appear as fundamental elements in I.

A more convenient point z is obtained by adding the points of intersection of t with the director spaces. Each z_i is of order $(m_1 + m_2 - (n-2))$ in ϕ_1, ϕ_2 and contains y to order $n-2$ apart from ϕ. Hence the order of the transformation I is

$$4k(m_1 + m_2 - n + 2) + 2(n-2) + 3.$$

Let $p = k(m_1 + m_2 - (n-2))$. Then the order of I is $4p + 2n - 1$. The base C_4 appears to multiplicity $2p+1$. The manifold γ of $n-2$ dimensions is double; the image of each point on it is a conic. By eliminating λ between (2) and (3) and making use of (1), the order of γ is found to be $r = 2p + 2k(n-2) + n - 1$. The order of Σ, locus of the parasitic conics may be obtained by determining the plane associated with each point

[12] Snyder and Rusk, "The Veneroni transformation in S_n," *Bulletin of the American Mathematical Society*, vol. 42 (1936), pp. 585-592.

of intersection of two director spaces, and getting the k conics cut from the associated quadric primals by this plane. This set of conics, as the point describes the intersection of the director spaces, describes Σ. It is of order $(n-2)(4p+n-1)-r$. The manifold of bisecants of C_4 and multisecants of γ is of order $4p+6n-10$. Since each quadric primal of the pencil goes into itself by I, the order of K is at once found to be $8p+4n-4$, hence that of Γ is $(n-1)(4p+2n-2)$. The table of characteristics for primes and the fundamental elements is now complete.

$$
\begin{aligned}
S_1 &\sim S_{4p+2n-1} : C_4^{2p+1}\, \gamma_r^{2}\, \Sigma^2 T \\
C_4 &\sim K_{8p+4n-2} : C_4^{4p+1}\, \gamma_r^{4}\, \Sigma^2 T^2 \\
\gamma &\sim \Gamma_{(n-1)(4p+2n-2)} : C_4^{4p+2}\, \gamma_r^{3}\, \Sigma^4 T^2 \\
\sigma &\sim \sigma_{2p+n+1} : C_4^{p+1}\quad \gamma_r \qquad \Sigma\, T \\
J &= K\Gamma.
\end{aligned}
$$

CORNELL UNIVERSITY,
AND
WELLS COLLEGE.

THE VENERONI TRANSFORMATION IN S_n.*

By Gertrude Blanch.

1. Introduction. The primals of S_n, of order n, which pass through $(n+1)$ general linear spaces S_{n-2} form a homaloidal system. The Cremona transformation of order n, determined by this system is known as the *Veroneroni*[1] transformation. Studies of it have been made by Eiesland[2] and more recently, by J. A. Todd[3] and by Virgil Snyder.[4] The author[5] studied the transformation in S_4; but the form used does not permit of a ready generalization for S_n. One general form has been obtained by Snyder and Rusk.[6] The form derived in this paper may be of interest because of its special symmetry. In addition, it is shown that in S_n, $n > 3$, it is necessary to satisfy $n(n-3)/2-1$ conditions in order that the transformation shall be involutorial.

2. Following C. A. Rupp,[7] let the linear forms S_{n-2} be defined by:

$$\phi_i:\quad x_i = 0;\quad A_i(x) = \begin{cases} \sum_{j=1}^{n+1} a_{ij}x_j = 0; & (i = 1, 2, \cdots, (n+1) \\ a_{ii} = 0. & \end{cases}$$

Consider the determinant of order $(n+1)$ given below:

$$(1)\qquad \begin{vmatrix} k_1 & a_{12}x_2 & a_{13}x_3 & \cdots & a_{1r}x_r & \cdots & a_{1,n+1}x_{n+1} \\ k_2 & -A_2(x) & a_{23}x_3 & \cdots & a_{2r}x_r & \cdots & a_{2,n+1}x_{n+1} \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdots & \cdot \\ k_r & a_{r2}x_2 & a_{r3}x_3 & \cdots & -A_r(x) & \cdots & a_{r,n+1}x_{n+1} \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdots & \cdot \\ k_{n+1} & a_{n+1,2}x_2 & a_{n+1,3}x_3 & \cdots & a_{n+1,r}x_r & \cdots & -A_{n+1}(x) \end{vmatrix} = 0.$$

* Received August 19, 1936; revised May 3, 1937.

[1] *Lombardo Rendiconti* II, vol. 34 (1901), pp. 640-654.

[2] *Palermo Rendiconti*, vol. 54 (1930), pp. 335-365.

[3] *Proceedings of the Cambridge Philosophical Society*, vol. 26 (1930), pp. 323-333.

[4] *Bulletin of the American Mathematical Society*, vol. 40 (1934), pp. 673-687.

[5] *American Journal of Mathematics*, vol. 58 (1936), pp. 639-645.

[6] *Bulletin of the American Mathematical Society*, vol. 42 (1936), pp. 585-592.

[7] *Bulletin of the American Mathematical Society*, vol. 35 (1929), pp. 319-320.

The cofactor of every k_j is satisfied by $x_i = A_i(x) = 0$ $(i = 2, 3 \cdots n+1)$. Moreover, the cofactor of every k_j is also satisfied by $x_1 = A_1(x) = 0$. For add to the second column the elements of all subsequent columns. The elements of the second column then become, respectively,

$$A_1(x), -a_{21}x_1, \cdots, -a_{r1}x_1, \text{ etc.}$$

It can be readily verified that the cofactor of k_i is $x_i W_i$, where W_i is a V_{n-1}^{n-1}, containing all the fundamental S_{n-2} except ϕ_i. Hence the determinant (1) is the complete linear system $| V_n{}^n |$, with k_i arbitrary. Let

$$x'_i = x_i W_i \qquad (i = 1, 2, \cdots, n+1). \tag{2}$$

Replacing the first column of (1) by the i-th column, we get a determinant with two columns alike, and on account of (2), we have the following identities:

$$M_i: \quad x_i \sum_j a_{ji} x'_j - A_i(x) x'_i = 0 \qquad (i = 2, \cdots, n+1).$$

Adding $M_2 + M_3 + \cdots M_{n+1}$, and remembering that $A_i(x) = \sum_j a_{ij} x_j$, $a_{ii} = 0$, we get

$$M_1: \quad x_1 \sum_j a_{j1} x'_1 - A_1(x) x'_1 = 0.$$

We therefore have the $(n+1)$ bilinear equations:

$$x_i \sum_j a_{ji} x'_j - A_i(x) x'_i = 0 \qquad (i = 1 \cdots n+1). \tag{3}$$

Write the matrix of the coefficients $\| a_{ij} \|$. It should be observed that the elements of the i-th row of this matrix determine the fundamental ϕ_i in (x) and similarly, the transpose of this same matrix gives at once the form of the fundamental S_{n-2} of (x'), namely:

$$\phi'_i: \quad x'_i = \sum_j a_{ji} x'_j = 0 \qquad (i = 1 \cdots n+1). \tag{4}$$

3. Involutorial transformations. If $a_{ij} = a_{ji}$ for every i, j, and (x), (x') are regarded as superposed, we have the identity transformation; if $a_{ij} = -a_{ji}$ for every i, j, then the transformation becomes a polarity among $(n+1)$ quadric primals, and is necessarily involutorial. However, by subjecting (x') to a linear transformation—equivalent to changing the frame of reference—we may obtain a form wherein the fundamental S_{n-2} in (x') coincide with their *associated*[8] S_{n-2} in (x), with fewer restrictions.

[8] By *associated* forms we mean those of (x) and (x') associated in the same bilinear equation.

To determine a suitable transformation carrying $\phi'_i \sim \phi_i$, we note from (4) that x'_i will have to be some linear combination of the two primes determining ϕ_i. Hence the transformation must be of the form

$$(5) \qquad x'_i = r_i y_i + s_i A_i(y), \qquad (i = 1 \cdots n+1).$$

As a consequence of (5)

$$(6) \qquad \sum_j a_{ji} x'_j \sim \sum_j a_{ji}(r_j y_j + s_j \sum_{k=1}^{n+1} a_{jk} y_k) = \mu_i(y).$$

It is necessary that $\mu_i(y)$ also be of the form

$$(7) \qquad \rho_i y_i + t_i A_i(y).$$

From (6) and (7)

$$(8) \qquad (\rho_i - \sum_j s_j a_{ji}^2) y_i + \sum_k (t_i a_{ik} - a_{ki} r_k - \sum_j s_j a_{ji} a_{jk}) y_k \equiv 0, \quad k \neq i.$$

The coefficient of every y_j must vanish. Consequently set $\rho_i = \sum_j s_j a_{ji}^2$. As to the second group of terms, we note that when $j = k$, $t_i a_{ik} - r_k a_{ki} = 0$, since $a_{kk} = 0$. Set $i = 1$, $k = 2$; and again $i = 2$, $k = 1$. We thus get

$$t_1 a_{12} - a_{21} r_2 = 0; \qquad t_2 a_{21} - a_{12} r_1 = 0.$$

Or, transposing,

$$a_{12}(t_1 + r_1) = a_{21}(t_2 + r_2).$$

Similarly, we must satisfy, in general

$$a_{ij}(t_i + r_i) = a_{ji}(t_j + r_j); \text{ for } (i, j = 1 \cdots (n+1).$$

Since the object is to impose as few restrictions as possible on the coefficients a_{ij}, set

$$(9) \qquad t_i = -r_i; \qquad (i = 1 \cdots (n+1).$$

From (8) and (9)

$$(10) \qquad r_i a_{ik} + r_k a_{ki} + \sum_j s_j a_{ji} a_{jk} = 0; \qquad (i, k = 1 \cdots (n+1); k \neq i).$$

By assigning all possible values to i and k, we obtain $n(n+1)/2$ linear equations in the $2(n+1)$ variables r_k, s_k. A study of the forms in S_5 and generalization to S_n shows that these equations determine $n(n-3)/2 - 1$ functionally independent determinants in the space of the coefficients a_{ij}, which must have the value zero if the conditions (10) are to be satisfied.[9]

[9] Thus in S_5 the 15×12 matrix gives 4 determinants of order 12, all having the last eleven rows of the matrix in common. The matrix of the last eleven rows has rank

When (10) holds, the forms are necessarily involutorial. For then the bilinear equations (3) are transformed into:

$$(11) \qquad x_i(\rho_i y_i - r_i \sum_j a_{ij} y_j) - \sum_j a_{ij} x_j \cdot (r_i y_i + s_i \sum_k a_{ik} y_k) = 0.$$

It is readily verified that the coefficient of $x_p y_q$ is the same as that of $x_q y_p$ for all subscripts. Hence the theorem: *The Veneroni Transformation in S_n can be made involutorial, with coincident associated fundamental elements, by imposing $n(n-3)/2-1$ conditions on the coefficients.* This verifies that in S_3 the transformation may always be made involutorial, and indeed in ∞' ways; in S_4, one condition must be imposed on the coefficients—as the author verified in the paper referred to. In S_5, four conditions among the coefficients of the form are necessary.

These involutorial transformations are not the only ones obtainable—the S_{n-2} of (x') need not all be associated with the same spaces of (x) in the bilinear equations. But the author's study in S_4 previously mentioned shows that other such involutorial transformations are always more specialized than the type where the S_{n-2} of (x') are the same as their associated S_{n-2} of (x).

HUNTER COLLEGE,
NEW YORK.

11; hence by well-known theorems about such matrices, the vanishing of these 4 determinants insures the vanishing of every other determinant of order 12. It is further possible to find sets of values of the coefficients a_{ij} for which the Jacobian of the four determinants with respect to certain four of the coefficients a_{ij} does not vanish. Moreover, for the most general values of these coefficients, the four determinants have no common factor. Consequently the 4 conditions are independent in S_5. The form of the determinants in S_5 permits a ready extension by induction to the independence of the conditions in S_n.

ON $2n$ POINTS WITH A REAL CROSS-RATIO.*

By F. MORLEY and J. R. MUSSELMAN.*

1. The viewpoint of Klein, that elementary geometry is not a study of isolated theorems but the theory of invariants of a group, is applied here to the inversive group and we study the meaning of an invariant of that group. Let us choose any $2n$ points $a_i (i = 1, 2, \cdots, 2n)$ in a plane (or on a sphere) which are ordered. Then, the expression

$$\frac{(a_1 - a_2)(a_3 - a_4) \cdots (a_{2n-1} - a_{2n})}{(a_2 - a_3)(a_4 - a_5) \cdots (a_{2n} - a_1)} = \rho, \tag{1.1}$$

where ρ is any real number, is invariant under all homographies

$$y = (\alpha x + \beta)/(\gamma x + \delta)$$

and under all antigraphies

$$\bar{y} = (\alpha x + \beta)/(\gamma x + \delta),$$

and hence is an invariant of the inversive group.

2. For $n = 2$, the cross-ratio

$$\frac{(a_1 - a_2)(a_3 - a_4)}{(a_2 - a_3)(a_4 - a_1)} = \rho \tag{2.1}$$

is the necessary and sufficient condition that *the four points a_i lie on a circle.*

For $n = 3$, we have six points a_i and

$$\frac{(a_1 - a_2)(a_3 - a_4)(a_5 - a_6)}{(a_2 - a_3)(a_4 - a_5)(a_6 - a_1)} = \rho. \tag{2.2}$$

Consider the three arcs $a_1a_2a_3$, $a_3a_4a_5$, $a_5a_6a_1$ (or else the arcs $a_2a_3a_4$, $a_4a_5a_6$, $a_6a_1a_2$). Let $a_1a_2a_3$ and $a_3a_4a_5$ meet again at a point b_3. Then from (2.1) we have

$$\frac{(a_1 - a_2)(a_3 - b_3)}{(a_2 - a_3)(b_3 - a_1)} = \rho_1 \quad \text{and} \quad \frac{(a_3 - a_4)(a_5 - b_3)}{(a_4 - a_5)(b_3 - a_3)} = \rho_2.$$

* Received April 30, 1937.

Hence, by using (2.2) we have

$$\frac{(a_5-a_6)(a_1-b_3)}{(a_6-a_1)(b_3-a_5)}=\rho_3.$$

That is, the three arcs $a_1a_2a_3$, $a_3a_4a_5$, $a_5a_6a_1$ meet at the point b_3. And equally the three arcs $a_2a_3a_4$, $a_4a_5a_6$, $a_6a_1a_2$ meet at a point, say b_4. The single condition (2.2) implies both facts. Two arcs $a_1a_2a_3$, $a_3a_4a_5$ which meet again at b_3 have a resultant arc $a_1b_3a_5$. The arcs $a_1a_2a_3$, $a_3a_4a_5$ and $a_5b_3a_1$ are then termed *balanced*. Thus for an ordered hexagon the invariant (2.2) means that *the arcs* $a_1a_2a_3$, $a_3a_4a_5$ and $a_5a_6a_1$ (or $a_2a_3a_4$, $a_4a_5a_6$ and $a_6a_1a_2$) *are balanced.*

For $n=4$, we have eight points and

$$\frac{(a_1-a_2)(a_3-a_4)(a_5-a_6)(a_7-a_8)}{(a_2-a_3)(a_4-a_5)(a_6-a_7)(a_8-a_1)}=\rho. \tag{2.3}$$

Consider the four arcs $a_1a_2a_3$, $a_3a_4a_5$, $a_5a_6a_7$ and $a_7a_8a_1$ (or else the arcs $a_2a_3a_4$, $a_4a_5a_6$, $a_6a_7a_8$ and $a_8a_1a_2$). Let $a_1a_2a_3$ and $a_3a_4a_5$ meet again at b_3, and let $a_5a_6a_7$ and $a_7a_8a_1$ meet again at b_7. Then from (2.1) we have

$$\frac{(a_1-a_2)(a_3-b_3)}{(a_2-a_3)(b_3-a_1)}=\rho_1, \qquad \frac{(a_3-a_4)(a_5-b_3)}{(a_4-a_5)(b_3-a_3)}=\rho_2,$$

$$\frac{(a_5-a_6)(a_7-b_7)}{(a_6-a_7)(b_7-a_5)}=\rho_3, \qquad \frac{(a_7-a_8)(a_1-b_7)}{(a_8-a_1)(b_7-a_7)}=\rho_4.$$

Hence, by using (2.3), we have

$$\frac{(a_5-b_3)(a_1-b_7)}{(b_3-a_1)(b_7-a_5)}=\rho_5,$$

which implies that the points a_5, b_3, a_1 and b_7 are concyclic. Thus for an ordered octagon the invariant (2.3) is satisfied *if the four points* a_1, a_5, *the second meet of* $a_1a_2a_3$ *and* $a_3a_4a_5$, *and the second meet of* $a_5a_6a_7$ *and* $a_7a_8a_1$ *are concyclic.*

This can also be stated as follows: *if we choose any two points x and y concyclic with a_1 and a_5 such that the two sets of arcs $a_1a_2a_3$, $a_3a_4a_5$, a_5xa_1 and $a_5a_6a_7$, $a_7a_8a_1$, a_1ya_5 are balanced, the invariant* (2.3) *is satisfied.* For this implies that $a_5b_3a_1b_7$ are concyclic.

If we call the second intersection of the arcs $a_{i-2}a_{i-1}a_i$ and $a_ia_{i+1}a_{i+2}$ by b_i (all subscripts taken modulo 8) we have the theorem that *if the four points* a_5, b_3, a_1 and b_7 *are concyclic, so also are* a_6, b_4, a_2, b_8; a_7, b_5, a_3, b_1 *and* a_8, b_6, a_4, b_2.

For the condition from (2.1) that a_{i+1}, b_{i-1}, a_{i-3}, b_{i+3} be concyclic reduces to

$$(2.4) \qquad \frac{(a_{i-3}-a_{i-2})(a_{i-1}-a_i)(a_{i+1}-a_{i+2})(a_{i+3}-a_{i+4})}{(a_{i-2}-a_{i-1})(a_i-a_{i+1})(a_{i+2}-a_{i+3})(a_{i+4}-a_{i-3})}=\rho.$$

If we let $i=4$ or 6 in (2.4) we obtain (2.3); if we let $i=5$ or 7 in (2.4) we obtain the reciprocal of (2.3) which proves the theorem. Thus if it happens that one set of points a_{i+1}, b_{i-1}, a_{i-3}, b_{i+3} for an ordered octagon be concyclic, there are three other concyclic sets. The single condition (2.3) implies all four.

In the general case of $2n$ ordered points $a_1a_2a_3a_4\cdots a_{2n}$, we replace the arcs $a_1a_2a_3$, $a_3a_4a_5$ by the arc $a_1b_3a_5$ where b_3 is the other intersection of the two arcs. We have now $2n-2$ points

$$a_1b_3a_5a_6a_7\cdots.$$

We replace the arcs $a_1b_3a_5$, $a_5a_6a_7$ by the arc $a_1c_5a_7$ where c_5 is the second intersection of the two arcs $a_1b_3a_5$ and $a_5a_6a_7$. We have now $2n-4$ points

$$a_1c_5a_7a_8a_9a_{10}\cdots.$$

Proceeding in this manner we arrive at four points. If these are concyclic, then the original set (or any intermediate set) has a real cross-ratio. For we have taken b_3 such that

$$\frac{(a_1-a_2)(a_3-b_3)}{(a_2-a_3)(b_3-a_1)}=\rho_1 \quad \text{and} \quad \frac{(a_3-a_4)(a_5-b_3)}{(a_4-a_5)(b_3-a_3)}=\rho_2$$

or

$$\frac{(a_1-a_2)(a_3-a_4)}{(a_2-a_3)(a_4-a_5)}=\rho_1\rho_2 \quad \frac{a_1-b_3}{a_5-b_3}$$

so that instead of

$$\frac{(a_1-a_2)(a_3-a_4)(a_5-a_6)\cdots}{(a_2-a_3)(a_4-a_5)(a_6-a_7)\cdots}=\rho$$

we have

$$\frac{(a_1-b_3)(a_5-a_6)\cdots}{(b_3-a_5)(a_6-a_7)\cdots}=\rho'.$$

To sum up: if we have a closed arc polygon $a_1a_2a_3$, $a_3a_4a_5$, $\cdots$, $a_{2n-1}a_{2n}a_1$ it is balanced if upon replacing $a_1a_2a_3$, $a_3a_4a_5$ by the resultant arc $a_1b_3a_5$, the new polygon is balanced. That is, if by continuing the process we get four points on a circle. If we have such a polygon, then taking $a_2, a_4\cdots a_{2n}$

arbitrarily on their arcs we get $2n$ points with a real cross-ratio. In particular if a_2 approaches a_1, if a_4 approaches a_3, and so on, then

$$\frac{\delta a_1 \cdot \delta a_3 \cdot \cdot \cdot \delta_{2n-1}}{(a_1 - a_3)(a_3 - a_5) \cdot \cdot \cdot (a_{2n-1} - a_1)} \text{ is a real} = \delta\rho.$$

For $n = 6$ we can give another interpretation of

$$(2.5) \qquad \frac{(a_1 - a_2)(a_3 - a_4) \cdot \cdot \cdot (a_{11} - a_{12})}{(a_2 - a_3)(a_4 - a_5) \cdot \cdot \cdot (a_{12} - a_1)} = \rho.$$

Let the second intersection of arcs $a_{i-2}a_{i-1}a_i$ and $a_ia_{i+1}a_{i+2}$ be denoted by b_i. We shall derive the condition that the three circles $a_{i-6}b_{i-4}a_{i-2}$, $a_{i-2}b_ia_{i+2}$ and $a_{i+2}b_{i+4}a_{i-6}$ meet at a point. The condition that b_{i-4} is the second intersection of $a_{i-6}a_{i-5}a_{i-4}$ and $a_{i-4}a_{i-3}a_{i-2}$ is

$$\frac{(a_{i-6} - a_{i-5})(a_{i-4} - a_{i-3})(a_{i-2} - b_{i-4})}{(a_{i-5} - a_{i-4})(a_{i-3} - a_{i-2})(b_{i-4} - a_{i-6})} = \rho_1.$$

The condition that b_i is the second intersection of $a_{i-2}a_{i-2}a_i$ and $a_ia_{i+1}a_{i+2}$ is

$$\frac{(a_{i-2} - a_{i-1})(a_i - a_{i+1})(a_{i+2} - b_i)}{(a_{i-1} - a_i)(a_{i+1} - a_{i+2})(b_i - a_{i-2})} = \rho_2.$$

The condition that b_{i+4} is the second intersection of $a_{i+2}a_{i+3}a_{i+4}$ and $a_{i+4}a_{i+5}a_{i-6}$ is

$$\frac{(a_{i+2} - a_{i+3})(a_{i+4} - a_{i+5})(a_{i-6} - b_{i+4})}{(a_{i+3} - a_{i+4})(a_{i+5} - a_{i-6})(b_{i+4} - a_{i+2})} = \rho_3.$$

The condition that circles $a_{i-6}b_{i-4}a_{i-2}$, $a_{i-2}b_ia_{i+2}$, $a_{i+2}b_{i+4}a_{i-6}$ meet at a point is from (2.2)

$$\frac{(a_{i-6} - b_{i-4})(a_{i-2} - b_i)(a_{i+2} - b_{i+4})}{(b_{i-4} - a_{i-2})(b_i - a_{i+2})(b_{i+4} - a_{i-6})} = \rho_4.$$

The product of these four expressions is

$$\frac{(a_{i-6} - a_{i-5})(a_{i-4} - a_{i-3})(a_{i-2} - a_{i-1})(a_i - a_{i+1})(a_{i+2} - a_{i+3})(a_{i+4} - a_{i+5})}{(a_{i-5} - a_{i-4})(a_{i-3} - a_{i-2})(a_{i-1} - a_i)(a_{i+1} - a_{i+2})(a_{i+3} - a_{i+4})(a_{i+5} - a_{i-6})} = \rho.$$

For $i = 7$ or 9—subscripts are taken modulo 12—we get (2.5); for $i = 8$ or 10 we get the reciprocal of (2.5). Other values of i, due to the symmetry, repeat. Hence the invariant (2.5) for an ordered twelve-point means that *the circles* $a_1b_3a_5$, $a_5b_7a_9$, $a_9b_{11}a_1$ *are on a point.* Equally it means that circles $a_2b_4a_6$, $a_6b_8a_{10}$, $a_{10}b_{12}a_2$; $a_3b_5a_7$, $a_7b_9a_{11}$, $a_{11}b_1a_3$; $a_4b_6a_8$, $a_8b_{10}a_{12}$, $a_{12}b_2a_4$ are on a point. The single condition (2.5) implies all four facts.

In similar fashion one can prove for sixteen points that if we call the

second intersection of $a_{i-2}a_{i-1}a_i$ and $a_ia_{i+1}a_{i+2}$ by b_i and the second intersection of $a_{i-4}b_{i-2}a_i$ and $a_ib_{i+2}a_{i+4}$ by c_i then the reality of the cross-ratio (1.1) implies that $a_{i-6}c_{i-2}a_{i+2}c_{i+6}$ are concyclic. There are eight such sets of four points, such if one set is concyclic, all eight are. We list them

$a_1c_5a_9\ c_{13}$	$a_5c_9\ a_{13}c_1$
$a_2c_6a_{10}c_{14}$	$a_6c_{10}a_{14}c_2$
$a_3c_7a_{11}c_{15}$	$a_7c_{11}a_{15}c_3$
$a_4c_8a_{12}c_{16}$	$a_8c_{12}a_{16}c_4$

For ten points the invariant (1.1) implies that *the circles* $a_{i-4}b_{i-2}a_i$ *and* $a_ib_{i+2}a_{i+4}$ *have* a_{i+5} *as their second point of intersection.* If it happens once it happens for all ten pairs of circles.

For fourteen points, we can state that if the invariant (1.1) be satisfied then *circles* $a_{i-8}c_{i-4}a_i$ *and* $a_ib_{i+2}a_{i+4}$ *have* a_{i+5} *as their second point of intersection.* If it happens once, it happens for all fourteen sets of circles.

3. A similar argument applies when the cross-ratio of the $2n$ points is a pure imaginary ρi. The circle on abc can here be replaced by the Apollonian circle on b and about ac. Let us investigate six ordered points where

$$\frac{(a_1-a_2)(a_3-a_4)(a_5-a_6)}{(a_2-a_3)(a_4-a_5)(a_6-a_1)}=\rho i. \tag{3.1}$$

The Apollonian circle on a_2 and about a_1a_3 may be written

$$\frac{(a_1-a_2)(a_3-x)}{(a_3-a_2)(a_1-x)}=\rho_1 i.$$

Similarly the Apollonian circles on a_4 and about a_3a_5, on a_6 and about a_5a_1 are

$$\frac{(a_3-a_4)(a_5-x)}{(a_5-a_4)(a_3-x)}=\rho_2 i$$

$$\frac{(a_5-a_6)(a_1-x)}{(a_1-a_6)(a_5-x)}=\rho_3 i.$$

The product of these three expressions is (3.1). Hence for *an ordered hexagon the invariant* (3.1) *means that the Apollonian circles on* a_2 *and about* a_1a_3, *on* a_4 *and about* a_3a_5, *on* a_6 *and about* a_5a_1 *meet at a point.* Equally the Apollonian circles on a_3 and about a_2a_4, on a_5 and about a_4a_6, on a_1 and about a_6a_2 will meet at a point. The single condition (3.1) implies both facts.

4. We can express the condition that any n ordered points should lie

on a line in terms of the Lagrange resolvents[1] of the n-point. We term as Lagrange resolvents of an ordered n-point the $n-1$ expressions

$$V_i = a_1 + \epsilon a_2 + \epsilon^2 a_3 \cdots \epsilon^{n-1} a_n$$

where ϵ is a root of $\epsilon^n = 1$ other than $\epsilon = 1$. We can take the n points as $a_i = a + \rho_i b$, $(i = 1, \cdots, n)$ where the ρ_i are real. Consider for simplicity four points. Then

$$\begin{aligned} V_1 &= b[(\rho_1 - \rho_3) + i(\rho_2 - \rho_4)] \\ V_2 &= b[\rho_1 - \rho_2 + \rho_3 - \rho_4] \\ V_3 &= b[(\rho_1 - \rho_3) - i(\rho_2 - \rho_4)] \end{aligned}$$

whence

$$V_1\bar{V}_1 = V_3\bar{V}_3, \quad V_2\bar{V}_1 = V_3\bar{V}_2 \quad \text{and} \quad V_2\bar{V}_3 = V_1\bar{V}_2$$

or

$$V_1/\bar{V}_3 = V_2/\bar{V}_2 = V_3/\bar{V}_1.$$

And so in general, for n ordered points to be on a line, we have

$$V_1/\bar{V}_{n-1} = V_2/\bar{V}_{n-2} = \cdots = V_{n-1}/\bar{V}_1.$$

THE JOHNS HOPKINS UNIVERSITY,
AND
WESTERN RESERVE UNIVERSITY.

[1] The Morleys, *Inversive Geometry*, p. 203.

EXTRACT FROM A LETTER BY E. CARTAN CONCERNING MY NOTE: ON CLOSED SPACES OF CONSTANT MEAN CURVATURE.*

By T. Y. Thomas.

In the proof of the theorem in my recent note in the *American Journal of Mathematics* (Vol. LVIII, 1936, p. 702) on hypersurfaces of a Euclidean space the assumption was made that the hypersurfaces were closed (but without boundary). I am in receipt of a letter from Professor E. Cartan following a mathematical discussion at Princeton in which he shows that this theorem can be proved without the above assumption, i. e. the theorem is in reality of local character. Cartan's proof forms an interesting supplement to my previous note. I present it (with permission) as an extract taken verbatim from his letter:

"En ce qui concerne votre intéressant théorème, dont vous m'aviez du reste parlé à Princeton, sur les espaces à courbure moyenne constante, je ne sais pas si je trompe, mais il me semble qu'il se ramène à une propriété purement locale et qu'il n'est pas necéssaire de supposer l'espace clos. Si j'ai bien compris, il s'agit de démontrer qu'une hypersurface S à $n \geqq 3$ dimensions de l'espace euclidean à $n+1$ dimensions est à courbure constants si, considérée comme espace riemannien elle satisfait aux conditions $B_{\alpha\beta} = \lambda g_{\alpha\beta}$ avec $\lambda > 0$.

"Soient en effet $a_1, a_2, \cdots, a_n$ les courbures principales de l'hypersurface S. Les conditions imposées sont

$$\begin{aligned} a_1a_2 + a_1a_3 + \cdots + a_1a_n &= \lambda, \\ a_2a_1 + a_2a_3 + \cdots + a_2a_n &= \lambda, \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad & \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad & \cdot \quad \cdot \\ a_na_1 + a_na_2 + \cdots + a_na_n &= \lambda, \end{aligned}$$

on encore, un désignant par S les somme des courbures,

$$a_i^2 - a_iS + \lambda = 0.$$

"Toutes les courbures principales satisfont donc à une même equation du second degré

$$x^2 - xS + \lambda = 0.$$

"Supposons qu'elles ne soient pas toutes égales entre elles; et que par exemple $a_1 \neq a_2$. Toutes les autres courbures seront égales soit à a_1, soit à a_2. Supposons que a_1 se présente p fois, et a_2 q fois $(p + q = n)$. On aura alors

* Received July 12, 1937.

$$a_1 + a_2 = S, \qquad a_1 a_2 = \lambda,$$

et, par suite,

$$(p-1)a_1 + (q-1)a_2 = 0, \qquad a_1 a_2 > 0.$$

Ces deux conditions sont contradictoires.

"Toutes les courbures principales étant égales, l'hypersurface est une hypersphere de rayon $1/|a_1|$."

It suffices for the above demonstration that the hypersurface S is defined by functions $\phi^i(x)$ where $i = 1, \cdots, n+1$ which are continuous and possess continuous partial derivatives to the third order (i. e. S is a hypersurface of class C^3). Including the case $n = 2$ for which the above result is immediately valid we may state the following theorem:

Any hypersurface of class C^3 of constant mean curvature $\lambda > 0$ and dimensionality $n \geqq 2$ in a Euclidean space of $n+1$ dimensions is a space of constant curvature.

Remark. One can find n mutually orthogonal directions $\lambda_1, \cdots, \lambda_n$ at a point P of the hypersurface, these directions being given as solutions of the equations $\Sigma_j(b_{ij} - a_k g_{ij})\lambda_k{}^j = 0$ where the b_{ij} are the coefficients of the second fundamental form and the a_k are the (real) roots of the determinant equation $|b_{ij} - a g_{ij}| = 0$. The $\lambda_1, \cdots, \lambda_n$ define the directions of lines of curvature at P and the $a_1, \cdots, a_n$ are (by definition) the normal curvatures at P of the corresponding lines of curvature. Denote by r_{ij} the Gaussian curvature at P of the geodesic surface determined by the directions λ_i and λ_j $(i \neq j)$. Then $r_{ij} = a_i a_j$ (Eisenhart, *Riemannian Geometry*, 1926, p. 198). Now $\Sigma_j r_{ij} = \lambda$ $(i \neq j)$ as a consequence of the condition that the hypersurface be of constant mean curvature λ (Eisenhart, *loc. cit.*, p. 113). Substituting $r_{ij} = a_i a_j$ in these latter equations we obtain the above equations of Cartan from which it follows that $a_j = \cdots = a_n = a$. Hence from

$$\Sigma_j(b_{ij} - a g_{ij})\lambda_k{}^j = 0 \text{ we have } b_{ij} = a g_{ij} \text{ at } P.$$

Then from the Gauss equations for the hypersurface

$$B_{\alpha\beta\gamma\delta} = b_{\alpha\delta} b_{\beta\gamma} - b_{\alpha\gamma} b_{\beta\delta}$$

we have

$$B_{\alpha\beta\gamma\delta} = a^2 (g_{\alpha\delta} g_{\beta\gamma} - g_{\alpha\gamma} g_{\beta\delta}) \text{ at } P.$$

These equations mean that the hypersurface is of constant curvature provided that the factor a^2 is a constant, i. e. independent of the point P. But multiplying the equations through by $g^{\alpha\delta} g^{\beta\gamma}$ and summing on repeated indices we obtain $\lambda = a^2(n-1)$ from which it follows that a^2 is a constant as required.

PRINCETON UNIVERSITY.

ALMOST PERIODIC FUNCTIONS AND HILL'S THEORY OF LUNAR PERIGEE.*

By AUREL WINTNER.

From the analytical point of view, Hill's theory of the mean motion of the perigee depends on three problems:

(i) determination of the intermediary periodic path of the Moon;

(ii) determination of the characteristic exponent of the orbit thus obtained;

(iii) identification of this characteristic exponent with the mean motion of lunar perigee.

Inequalities which are due to the eccentricities are not considered in Hill's theory and have been developed by Brown. The same holds for the parallactic terms, Hill's equations of motion being derived by a modification of the restricted problem of three bodies.

Hill[1] solved both problems (i), (ii) by a bold application of infinitely many variables, (ii) leading to a certain linear Eigenwertproblem of specific type, and (i) to a non-linear and non-recursive system of infinitely many equations for infinitely many Fourier constants. Hill's application of infinite determinants in the linear problem (ii) has been justified by Poincaré.[2] The question of existence and convergence in case of the non-linear system arising in problem (i) is of quite another nature. Hill was well aware of the novelty of this analytical question and he formulated its mathematical treatment as a desideratum.[3] The proofs requested by Hill were supplied some years ago.[4]

* Received June 20, 1937.

[1] G. W. Hill, *Collected Mathematical Works*, vol. 1 (1905), pp. 284-335 and pp. 243-270. Cf. also the presentation of Poincaré in vol. 2, part 2 (1909), chap. XXV-XXVII of his *Leçons de Mécanique Céleste.*

[2] Cf., e. g., H. Poincaré, *loc. cit.* [1], pp. 49-56.

[3] G. W. Hill, *loc. cit.* [1], p. 287. Poincaré's presentation (*loc. cit.* [1], pp. 31-40) of (i) does not satisfy Hill's desideratum; the legalization of Hill's method in (i) is of a later date.[4] It can be mentioned that the question of the existence of a family of simple closed orbits about the Earth is not identical with the question whether Hill's method of infinitely many variables in the non-linear problem (i) (i. e., the method actually applied in lunar theory) is or is not a procedure which has mathematical legality.

Thus the mathematical problems arising in (i) and (ii) are solved. The present note deals with the mathematical problem in (iii) which presupposes (i) and (ii) in the same way as (ii) presupposes (i).

Hill's justification of the step (iii) is not of a mathematical nature, since it merely consists [5] of an interpretation of Hill's results in terms of the formal theory of Delaunay. Since the Hill-Brown theory is, in the main, independent of the older lunar theories, it seems desirable to give a direct account of the step (iii). From the mathematical point of view, the objection to the usual explanation [5] of (iii) is not merely a matter of method. In fact, it is a well-defined question, whether or not there exists at all a mean motion, and if it does, whether or not its remainder term shows a recurrent behavior which, as tacitly assumed in lunar theory, can be analyzed into an anharmonic Fourier series.

In what follows, these questions will be answered in the affirmative. The proof depends on an appropriate application of a theorem on almost periodic functions which was formulated as a conjecture by the present author and subsequently proved by Bohr.[6] This theorem of Bohr states that if $\zeta(t)$ is a complex-valued almost periodic function of the real variable t and $|\zeta(t)| > \alpha$ holds for some constant $\alpha > 0$ and for every t, then the real continuous function $\arg \zeta(t)$ can be decomposed into the sum of a linear "secular" term μt and of a "recurrent" term, the latter term being almost periodic, while μ is a constant. Here and later on, almost periodicity is meant in the original sense of Bohr.

The problem concerns certain of the solutions of the three Jacobi equations for isoenergetic variations, this system of differential equations being represented by the equation of second order for normal displacements and by the equation of first order for tangential displacements.[7] While the latter

---

According to G. D. Birkhoff ["The Restricted Problem of Three Bodies," *Rendiconti del Circolo Matematico di Palermo*, vol. 39 (1915), part 3, art. 12], the existence of the simple closed orbits in question can be proved in a rather easy way. While such an existence proof was claimed by Poincaré (*loc. cit.* [1], p. 30), it seems to the writer that the passage of the *Méthodes Nouvelles* which is referred to by Poincaré does not imply such a proof, a proof being essentially dependent on Birkhoff's λ-device (cf. H. Poincaré, *loc. cit.* [1], p. 30, line 6) or on some equivalent substitution (facilitating expansions which are *uniformly* convergent in the vicinity of the mass singularity).

[4] A. Wintner, "Zur Hillschen Theorie der Variation des Mondes," *Mathematische Zeitschrift*, vol. 24 (1925), pp. 259-265; also "Ueber die Konvergenzfragen der Mondtheorie," *ibid.*, vol. 30 (1929), pp. 211-227.

[5] G. W. Hill, *loc. cit.* [1], pp. 269-270; also H. Poincaré, *loc. cit.* [1], p. 67.

[6] H. Bohr, "Kleinere Beiträge zur Theorie der fastperiodischen Funktionen, I," *Det Kgl. Danske Videnskabernes Selskab Math.-fys. Meddelelser*, vol. 10, no. 10 (1930).

[7] Cf. G. D. Birkhoff, *loc. cit.* [3], part 1, art. 3.

equation is not needed for the determination of the characteristic exponents, the geometrical problem of the mean motion of lunar perigee cannot be treated by considering the normal displacements alone, instead of the whole problem in the (x, y)-plane. Correspondingly, the situation is different from the one arising in the Adams theory of the mean motion of the lunar node, a problem which is defined in terms of a linear canonical periodic system with a *single* degree of freedom and is, therefore, reducible to the torus problem of Poincaré-Denjoy.[8] In other words,[9] the result to be obtained can be applied to, but is not identical with, an anharmonic Fourier analysis of a sequence of conjugate points on the closed extremal of problem (i), the rotation number being represented by the coefficient μ of the almost periodic remainder term $\arg \zeta(t) - \mu t$.

Hill's equations of motion are of the form

$$(1) \qquad x'' - 2y' = \Omega_x(x, y), \qquad y'' + 2x' = \Omega_y(x, y),$$

where the primes indicate derivatives with respect to t. Placing

$$(2) \qquad t = m\tau,$$

where m is Hill's parameter determining a solution of (1) within his family of periodic orbits, these periodic solutions are of the form [1]

$$(3) \quad x = \sum_{k=-\infty}^{+\infty} a_k \cos(2k+1)\tau, \quad y = \sum_{k=-\infty}^{+\infty} a_k \sin(2k+1)\tau, \quad \text{where} \quad a_k = a_k(m).$$

The Fourier constants $a_k(m)$ are analytic functions of m and vanish at $m = 0$ with increasing rapidity as $k \rightarrow \pm\infty$. Hill's procedure (i) leading to (3) has been justified [4] for $|m| < 0.08333 \cdots$, and so, in particular, for the case $m = +0.08084 \cdots$ of the Moon.

In problem (ii), the validity of Hill's method does not depend on the magnitude of m, since this problem is linear. Exceptional is only the case of an m for which the path (3) has a point on its curve of zero velocity (or reaches the mass singularity). In fact, for this case the equation of normal displacement [10] acquires a singularity, and no direct procedure seems to have

[8] T. Levi-Civita, "Sur les équations linéaires à coefficients périodiques et sur le moyen mouvement du noeud lunaire," *Annales de l'Ecole Normale Supérieure*, ser. 3, vol. 28 (1911), pp. 325-376.

[9] Cf. G. D. Birkhoff, "Dynamical systems with two degrees of freedom," *Transactions of the American Mathematical Society*, vol. 18 (1917), part 2, art. 14, and the investigations of Poincaré referred to there.

[10] As to this equation, cf. A. Wintner, "Ueber die Jacobische Differentialgleichung des restringierten Dreikörperproblems," *Berichte der Sächsischen Akademie der Wissenschaften*, vol. 82 (1930), pp. 345-354.

been developed for the resulting singular Eigenwertproblem. The first positive[11] singular value of m is, however, quite large and is represented by the value $m = 0.56 \cdots$ belonging to Hill's cuspidal orbit of maximum lunation,[12] a value beyond the m-range under consideration.

The Jacobi equations belonging to (1) and (3) can be written as

$$(4) \qquad \begin{aligned} \xi'' - 2\eta' &= \Omega_{xx}(t; m)\xi + \Omega_{xy}(t; m)\eta, \\ \eta'' + 2\xi' &= \Omega_{xy}(t; m)\xi + \Omega_{yy}(t; m)\eta, \end{aligned}$$

and have, for fixed m, four linearly independent solutions $(\xi(t), \eta(t))$ two of which are of a trivial type, while the remaining two, which represent isoenergetic displacements, supply the general solution of the equation of (isoenergetic) normal displacements; cf. (ii). The pair of trivial solutions of (4) is obtained by differentiating the solution (3) of (1), where $\tau = t/m$, with respect to m and the origin of the t-axis respectively, so that these solutions of (4) belong to the trivial pair $(0, 0)$ of characteristic exponents and the first of these two solutions contains a secular term and is not an isoenergetic displacement. Since the orbit (3) is symmetrical with respect to both coördinate axes $y = 0$, $x = 0$ and is formally stable in the m-domain under consideration, it is readily inferred from the Fuchs-Floquet theory that, if $(c, -c)$ represents the pair of non-trivial characteristic exponents, the two non-trivial solutions of (4) can be represented in the form

$$(5) \qquad \begin{aligned} \xi(t; m) &= \gamma \sum_{k=-\infty}^{+\infty} A_k \cos\{(2k + 1 + c)\tau + \delta\}, \\ \eta(t; m) &= \gamma \sum_{k=-\infty}^{+\infty} B_k \sin\{(2k + 1 + c)\tau + \delta\}, \end{aligned}$$

---

[11] As to the corresponding range of the family (3) in the retrograde case $m < 0$, cf. G. D. Birkhoff, *loc. cit.*[3], part 4, art. 16.

[12] While Hill[1] had to derive this orbit, not from his analytical expansions, but by using mechanical quadratures, a mathematical existence proof for this orbit can be considered as implied by Strömgren's general termination principle, the existence of the family (3) being established for small $m > 0$. Unfortunately, sufficient material is not available concerning the natural termination of Hill's family, and the same holds precisely for that among the Copenhagen groups which corresponds to Hill's family, i. e., for Strömgren's Klasse g. The analytic continuation of the lunar orbits (3) which is given by Kelvin in vol. 4, no. 55, of his *Collected Works,* is rather flat with fully developed loops (the latter arising at the stage of Hill's cuspidal orbit). Judging from Kelvin's single orbit, which may or may not be sufficiently advanced, one possibility for the natural termination would be an indefinite flattening in the direction perpendicular to the axis of syzygies, the orbit becoming the whole y-axis, while the Sun is infinitely distant and on the x-axis.

where $\gamma \neq 0$ and δ are arbitrary real integration constants and

$$c = c(m); \qquad A_k = A_k(m), \qquad B_k = B_k(m) \qquad\qquad (m = t/\tau)$$

are real analytic functions of m.

As $m \to 0$, the behavior of $A_k(m)$ and $B_k(m)$ in (5) is, for large $|k|$, similar to that of $a_k(m)$ in (3). In particular, the limiting values $(A_k)_{m=0}$ and $(B_k)_{m=0}$ vanish unless $|2k+1| = 1$. If $|2k+1| = 1$, so that $k = 0$ or $k = -1$, it is found from the expansions of $A_k(m)$ and $B_k(m)$ that

$$(6_1) \qquad A_0^0 \neq 0, \quad B_0^0 = A_0^0; \qquad A_{-1}^0 = -3A_0^0, \quad B_{-1}^0 = 3B_0^0,$$

where $A_k^0 = (A_k)_{m=0}$, $B_k^0 = (B_k)_{m=0}$. The limit relations (6_1), together with

$$(6_2) \qquad A_k^0 = 0, \quad B_k^0 = 0, \quad \text{where} \quad 2k+1 \neq \pm 1,$$

have been checked by comparison with Kepler's motion.

Letting $m \to 0$ in (5), using (6_1)–(6_2) and omitting the multiplicative integration constant $\gamma \neq 0$ (or, rather, $2A_0^0\gamma \neq 0$), one finds after an easy reduction that, if c^0 denotes the limiting value $(c)_{m=0}$ of the characteristic exponent $c = c(m)$,

$$(\xi)_{m=0} = -\cos\tau\cos(c^0\tau + \delta) - 2\sin\tau\sin(c^0\tau + \delta),$$
$$(\eta)_{m=0} = -\sin\tau\cos(c^0\tau + \delta) + 2\cos\tau\sin(c^0\tau + \delta).$$

Hence

$$(\xi^2 + \eta^2)_{m=0} = \cos^2(c^0\tau + \delta) + 4\sin^2(c^0\tau + \delta).$$

Since this continuous and non-vanishing function of τ is periodic, there exists a sufficiently small constant α_0 which depends on the integration constants δ and $\gamma \neq 0$ but is such that

$$(\xi^2 + \eta^2)_{m=0} > \alpha_0 > 0 \quad \text{for} \quad -\infty < \tau < +\infty.$$

Hence there exists a sufficiently small $m^* > 0$ such that if m has any value between $-m^*$ and m^*, one can choose a positive constant $\alpha = \alpha_m$ for which the almost periodic functions (5) of τ or $t = m\tau$ satisfy the inequality

$$(7) \qquad [\xi(t;m)]^2 + [\eta(t;m)]^2 > \alpha_m > 0; \qquad -\infty < t < +\infty,$$

it being understood that α_m is, for fixed m, a function of the integration constants γ, δ occurring in (5), while m^* is independent of γ, δ.

As pointed out above, the coefficients $A_k(m)$ and $B_k(m)$ tend, for fixed

small values of m, rather strongly to zero as $k \to \pm \infty$. Correspondingly, it requires only quite rough numeral estimates to show that (7) is satisfied for the value of m which belongs to the Moon, i. e., that $m = 0.08084 \cdots$ is "sufficiently small" or less than m^*.

For a fixed m and for fixed values of the integration constants $\gamma \neq 0$, δ, introduce the perihelium ϖ, or rather perigee, as a function of t or τ by placing

$$(8) \qquad \xi = \rho \cos \varpi, \quad \eta = \rho \sin \varpi, \quad \text{where} \quad \rho = (\xi^2 + \eta^2)^{\frac{1}{2}} \geqq 0.$$

It is understood that, since $\rho \neq 0$ by (7), the value of $\varpi = \varpi(t)$ is determined for every t by continuity, if one assigns an initial normalizing condition, say $0 \leqq \varpi(0) < 2\pi$. Put

$$(9) \qquad \zeta(t) \equiv \zeta = \xi + i\eta = (\xi^2 + \eta^2)^{\frac{1}{2}} \exp i\varpi; \quad \text{cf. } (8).$$

Then $\zeta(t)$ is, in view of (5) and (2), an almost periodic function. Furthermore, (7) shows that $|\zeta(t)| > \alpha_m^{\frac{1}{2}} > 0$ for every t. Consequently, the theorem of Bohr [6] is applicable to $\zeta(t) = |\zeta| \exp i\varpi$. It follows that the perigee $\varpi = \varpi(t)$ is of the form

$$(10) \qquad \varpi(t) = \mu t + \text{an almost periodic function of } t,$$

where μ is a constant.

It is instructive to look upon (10) from the point of view of the fundamental mathematical difficulty of celestial mechanics, a difficulty represented by the problem of small integration divisors in the astronomical theories and by Birkhoff's unsolved stability problem of incommensurable rotation numbers in the corresponding mathematical theories. First, from (8),

$$(11) \qquad \varpi' \equiv \frac{d\varpi}{dt} = \frac{\eta'\xi - \xi'\eta}{\xi^2 + \eta^2}.$$

It follows, therefore, from (7), (5) and (2) that the derivative $\varpi'(t) \equiv \varpi'(m\tau)$ of the perigee is an almost periodic function. Now the integral of an almost periodic function is almost periodic whenever it is bounded. Hence, if μ denotes the constant term in the Fourier series of the almost periodic function (11), so that

$$(12) \qquad \mu = \lim_{T \to \infty} \frac{1}{T} \int_0^T \varpi'(t)\, dt,$$

then (10) can be inferred from (11) by integration if and only if $\varpi(t) - \mu t$ is a bounded function. On the other hand, the integral of an almost periodic function which has no constant term in its Fourier series is, according to Favard, sure to be almost periodic if there are no frequencies of arbitrarily small absolute value, i. e., if there are no "small divisors"; while otherwise one cannot say anything, as shown precisely by the classical astronomical problem. Now all that follows from (11), (7), (5) and (2) is that the frequencies of the almost periodic function $\varpi'(m\tau)$ or $\varpi'(m\tau) - \mu$ are contained in the double sequence

$$(13) \qquad \nu_1 c + \nu_2, \quad \text{where} \quad \nu_1, \nu_2 = 0, \pm 1, \pm 2, \cdots.$$

Since $c = c(m)$ is a continuous and non-constant function of m and can, therefore, be considered as irrational, the double sequence (13) is identical with the sequence of frequencies occurring in the classical problem of small divisors. Accordingly, Bohr's theorem on $\arg \zeta(t)$ plays, in the above proof of (10), a definite dynamical rôle: it assures that the classical small divisors are harmless in the present case [although the continuous function $c = c(m)$ in (13) attains values which are represented by arbitrarily irregular continued fractions, no c-set of measure zero being excluded]. This could be checked by a direct discussion of the Fourier constants of the almost periodic function (11).

Since all frequencies of the almost periodic function $\varpi'(m\tau)$ of $\tau = t/m$ are contained in (13), the Fourier series of the almost periodic function on the right of (10) is of the form

$$(14) \qquad \varpi(t) - \mu t \sim \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} C_{jk} \cos \left\{ \frac{c(m)j + k}{m} t + \delta_{jk} \right\}.$$

This agrees with a result of Bohr,[13] according to which both the frequencies and the secular constant of $\varpi(m\tau)$ must be of the form (13); so that

$$(15) \qquad \mu = (cr + s)/m, \qquad c = c(m),$$

where r and s are integers.

Hill did not consider the Fourier series (14) but rather the constant μ of the secular part of the perigee $\varpi(t)$; and he wrote,[14] corrrespondingly,

---

[13] H. Bohr, "Ueber fastperiodische ebene Bewegungen," *Commentarii Mathematici Helvetici*, vol. 4 (1932), pp. 51-64.

[14] G. W. Hill, *loc. cit.* [1], pp. 269-270. Hill's notations are not the same as those used above; the difference is explained below.

$\varpi(t) = \mu t$ and $\varpi'(t) = \mu$. It is, however, obvious from the context that what Hill actually had in mind were relations of the type (14), (10) and (12). While this assumption of Hill is now justified, the corresponding mathematical postulates in Brown's theory, where the higher powers of the lunar eccentricity are not neglected, seem to require an entirely different approach.[15]

While (8) refers the perigee $\varpi(t)$ to the synodical coördinate system, Hill's perigee $\omega(t)$ is referred to a siderial coördinate system. Since the first of these frames, when considered from the second, rotates with constant angular velocity, $\varpi(t) - \omega(t)$ is a linear function of t, and so the above results hold for $\omega(t)$ also. Since $c = c(m)$ and $\mu = \mu(m)$ are continuous functions of m, while r and s in (15) are integers, one can determine the actual values of r, s for small m by letting $m \rightarrow 0$ and comparing the result with the formulae of Kepler's motion. This, when combined with the reduction of $\varpi(t)$ to $\omega(t)$, gives as connection between the mean motion of lunar perigee and the characteristic exponent c precisely the relation stated by Hill.[5]

THE JOHNS HOPKINS UNIVERSITY.

[15] Cf. A. Wintner, *loc. cit.* [4], pp. 214-215.

SUR LES MOUVEMENTS DES SYSTÈMES DYNAMIQUES QUI ADMETTENT " L'INCOMPRESSIBILITÉ " DES DOMAINES.*

Par HEINRICH HILMY.

1. Considérons un système dynamique dont les mouvements sont définis par l'équation différentielle

$$dx_i/dt = X_i(x_1, \cdots, x_n) \qquad (i = 1, \cdots, n). \tag{1}$$

Nous interprétons la variable indépendante t comme le temps et les variables $x_1, \cdots, x_n$, comme les coordonnées du point p se mouvant dans l'espace euclidien E^n, à n-dimensions.

Supposons que les seconds membres des équations (1) sont des fonctions continues et qu'ils satisfont aux conditions de Lipschitz dans les points des sous-ensembles considérés E^n. Dans ce cas auront lieu les théorèmes généraux de l'existence et de l'unicité des solutions, et également le théorème de la dépendance continue des solutions des conditions initiales.

Les mouvements du système dynamique sont caractérisés par les trajectoires que décrivent dans E^n les points $p \subset E^n$. Nous désignons par $f(p, t)$ la position du point p dans le moment du temps t; $f(p, 0) = p$.

Si A est un ensemble contenu dans E^n le symbole $f(A, t)$ désigne l'ensemble $\{f(p, t)\}$ où $p \subset A$.

Le symbole $S(x, \epsilon)$ désigne le voisinage sphérique du point p au rayon ϵ.

L'ensemble M est dit positivement (négativement) invariant, si $f(p, t) \subset M$ pour toute valeur de $t \geq 0$ $(t \leq 0)$. L'ensemble M est dit invariant s'il est en même temps positivement et négativement invariant.

Le mouvement $f(p, t)$ est dit positivement (négativement) stable au sens de Lagrange si l'on peut indiquer une sphère S dont le centre serait le point p et le rayon sufisemment grand pour que $f(p, t) \subset S$ pour tout $t \geq 0$ $(t \leq 0)$.

Le mouvement $f(p, t)$ est dit positivement (négativement) stable au sens de Poisson si pour tous arbitraires $\epsilon > 0$ et $T > 0$ on peut indiquer des valeurs de t positives (négatives) qui satisfassent à l'inégalité $|t| > T$ et pour lesquels $\rho(p, f(p, t)) < \epsilon$ [$\rho(x, y)$ désigne la distance entre les points x et y dans E^n]. Le mouvement $f(p, t)$ est dit stable au sens de Poisson s'il est tel simultanement dans le sens positive et dans le sens négative.

Convenons de dire que le mouvement instable au sens de Poisson et en

* Received May 3, 1937.

même temps stable au sens de Lagrange est un mouvement asymptotique. Si $f(p, t)$ est un mouvement positivement (négativement) instable au sens de Poisson, il découle de la continuité des seconds membres de l'équation (1) que pour tout point p' de la trajectoire $f(p, t)$ on peut indiquer un voisinage sphérique suffisemment petit pour que le point p', ayant une fois quitté ce voisinage n'y revienne jamais le temps changeant dans le sens positif (négatif). Nous dirons que le nombre positif η égal à la borne supérieure des rayons de tous les voisinages sphériques du point p' dans lesquels il ne revient plus après l'avoir une fois quitté, le temps changeant dans le sens positif (négatif),— nous dirons que ce nombre est l'indice caractéristique du point p' pour le temps croissant (décroissant).

2. Nous dirons que dans l'ensemble invariant M a lieu l'incompressibilité des domaines si pour tout domaine relatif borné $G \subset M$ se vérifie l'énoncé suivant: pour aucune valeur de t le domaine $f(G, t)$ n'est contenu dans une partie vraie de G.

Théorème. *Si dans l'ensemble fermé invariant M a lieu "l'incompressibilité" des domaines, tous les points de cet ensemble, sauf l'ensemble des points de la première catégorie de R. Baire, appartiennent à la somme des ensembles des points situés sur les trajectoires de mouvements stables au sens de Poisson et de ceux des points situés sur les trajectoires des mouvements instables au sens de Lagrange dans les deux sens.*

Démontrons, pour commencer, le théorème pour le temps croissant.

Désignons par Q l'ensemble de tels points $p \subset M$ qui sont situés sur les trajectoires des mouvements asymptotiques pour $t \to +\infty$. Prenons une suite de nombres positifs croissant indéfiniment et uniformement

$$R_1, R_2, \cdots, R_n, \cdots$$

et désignons par

$$S_1, S_2, \cdots, S_n, \cdots$$

l'ensemble des points $p \subset M$ situés dans les sphères et sur leurs frontières aux rayons respectivement égaux à $R_1, R_2, \cdots, R_n, \cdots$ décrites autour du point arbitraire appartenant à l'ensemble M.

Soit Q^k un ensemble de tels points $p \subset Q \cdot S_k$ pour lesquels on a $f(p, t) \subset S_k$ pour toute valeur de $t \geq 0$; alors

$$Q = \sum_{i=1}^{\infty} Q^i. \tag{2}$$

Montrons que chacun des ensembles Q^k est un ensemble de la première catégorie de Baire par rapport à S_k et par conséquent par rapport à M également.

Donnons nous une suite de nombres positifs

$$\epsilon_1, \epsilon_2, \cdots, \epsilon_n, \cdots$$

convergeant uniformément à zéro. Désignons par Q^k_m l'ensemble des points $p \subset Q^k$ dont l'indice caractéristique est $\eta \geq \epsilon_m$. Il est évident que

$$Q^k = \sum_{i=1}^{\infty} Q^k_i. \tag{3}$$

Démontrons que pour k fixé et pour tout m l'ensemble Q^k_m n'est dense nulle part dans S_k. Admettons le contraire; il se trouverait alors un domaine relatif $G \subset S_k$, dans lequel l'ensemble de points appartenant à Q^k_m serait partout dense.

Donnons nous un nombre arbitraire $\tau > 0$; choisissons dans le domaine G le point $p' \subset Q^k_m$ et décrivons autour de ce point la sphère S^* au rayon $\epsilon < \frac{1}{2}\epsilon_m$ et suffisemment petit pour que les conditions $S^* \subset G$ et $f(S^*, \tau) \cdot S^* = 0$ soient vérifiées, ce qui est possible en vertu de la dependance continuelle des conditions initiales. Montrons à present que $f(S^*, t) \cdot S^* = 0$ pour tout $t \geq \tau$. S'il n'était pas ainsi, on pourrait choisir un point $q \subset S^*$ et un nombre $T > \tau$ tels que $f(q, T) \subset S^*$. Soit $\epsilon^* > 0$ suffisemment petit pour que $S(q, \epsilon^*) \subset S^*$. En vertu de la dépendance continue des trajectoires des conditions initiales on peut indiquer un nombre $\delta > 0$ suffisemment petit pour que

$$f(S(q, \delta), T) \subset S(f(q, T), \epsilon^*).$$

Mais $S(q, \delta)$ contient nécessairement le point $p^* \subset Q^k_m$, $f(p^*, \tau) \not\subset S^*$ et $f(p^*, T) \subset S^*$ par conséquent le point p^*, ayant quitté la sphère S^* y revienne encore, ce qui est contradictoire vu que l'indice caractéristique de p^* surpasse ou égale ϵ_m.

Considérons le domaine

$$\Gamma = \sum_{t \geq 0} f(S^*, t).$$

Le domaine Γ est positivement invariant et borné,

$$f(\Gamma, \tau) = \sum_{t \geq \tau} f(S^*, t)$$

étant un sous-ensemble de Γ vrai vu que $S^* \subset \Gamma$ et $f(\Gamma, \tau) \cdot S^* = 0$ ce qui contredit la supposition que " l'incompressibilité " des domaines a lieu dans M. La contradiction à laquelle nous arrivons montre que l'ensemble Q^k_m n'est

dense nulle part dans S_k. Mais alors l'ensemble Q^k, en vertu de (3) et l'ensemble Q en vertu de (3) seront des ensembles de la première catégorie de Baire par rapport à M.

En désignant par N l'ensemble des points $p \subset M$ qui sont situés sur des trajectoires asymptotiques pour $t \to -\infty$ et en répétant les considérations faites plus haut, démontrons que l'ensemble N sera un ensemble de la première catégorie de R. Baire par rapport à M. Mais alors l'ensemble $A = Q + N$ sera également un ensemble de la première catégorie de R. Baire par rapport à M, et le théorèm est démontré.

3. Le théorème démontré complète les résultats obtenus dans le mémoire connu de E. Hopf[1] qui traite de la généralisation du Wiederkehrsatz de H. Poincaré en élucidant la catégorie au sens de R. Baire des ensembles étudiés dans ce mémoire. Le théorème de E. Hopf est applicable aux systèmes dynamiques qui admettent l'invariant intégral de H. Poincaré. Le théorème que nous venons de démontrer est applicable à une classe plus large de systèmes dynamiques, vu que l'incompressibilité des domaines peut avoir lieu dans des systèmes dynamiques qui n'admettent pas d'invariant intégral. Construisons un exemple d'un système dynamique qui justifie cette assertion.

Considérons un segment de l'axe de x, de la longueur 1 aux coordonnées des extrémités $-\frac{1}{2}$ et $+\frac{1}{2}$ et construisons sur ce segment un ensemble formé et partout non dense que nous désignons par P.

Prenons une suite de nombres

$$\alpha_n = 1/4^n \qquad (n = 1, 2, \cdot \cdot \cdot).$$

Construisons sur le premier segment un autre segment de la longueur α_1, de manière que son centre coïncide avec le centre du premier segment et excluons tous les points intérieurs du second segment. Sur chacun des deux segments conservés construisons de nouveau des segments de la longueur α_2, de manière que les centres de ces segments coïncident avec les centres des segments qui les contiennent; excluons ensuite encore tous les points intérieurs des segments que nous venons de construire. Représentons nous ce procédé indéfini. Désignons par G l'ensemble de tous les points du segment $[-\frac{1}{2}, +\frac{1}{2}]$ qui peuvent être exclus par ce procédé et par P l'ensemble de tous les autres points.

Il est évident que l'ensemble P est un ensemble fermé et partout non dense, tandis que l'ensemble G est dense partout sur le premier segment.

De simples calculs nous montrent que

[1] E. Hopf, "Zwei Sätze über den wahrscheinlichen Verlauf der Bewegungen dynamischer Systeme," *Mathematische Annalen*, Bd. 103 (1930).

$$\text{mes}\, P = \tfrac{1}{2}, \qquad \text{mes}\, G = \tfrac{1}{2}.$$

L'ensemble des intervalles, exclus du segment $[-\tfrac{1}{2}, +\tfrac{1}{2}]$ lors de la construction de l'ensemble P est un ensemble dénombrable. Donc nous pouvons les énumérer. Soit

$$\Delta_1, \Delta_2, \cdots, \Delta_n, \cdots$$

la suite de ces intervalles. Désignons par a_n et par b_n les coordonnées respectives de l'extrémité droite et de l'extremité gauche de l'intervalle Δ_n. Définissons sur le segment $[-\tfrac{1}{2}, +\tfrac{1}{2}]$ la fonction $\Phi(x)$ comme il suit:

$$\Phi(x) = \begin{cases} 0 \quad , & \text{si } x \subset P, \\ x - a_n, & \text{si } a_n \leq x \leq a_n + \dfrac{b_n - a_n}{2}, \\ b_n - x, & \text{si } a_n + \dfrac{b_n - a_n}{2} \leq x \leq b_n. \end{cases}$$

La fonction $\Phi(x)$ est une fonction continue sur tout le segment $[-\tfrac{1}{2}, +\tfrac{1}{2}]$. De plus on peut constater par des simples calculs que pour deux points quelconques x_1 et x_2 du segment $[-\tfrac{1}{2}, +\tfrac{1}{2}]$ se vérifie l'inégalité:

$$|\,\Phi(x_1) - \Phi(x_2)\,| \leq |\,x_1 - x_2\,|$$

c'est à dire que la fonction $\Phi(x)$ satisfait à la condition de Lipschitz.

Examinons dans le plan euclidean E^2 l'ensemble de points M:

$$-\tfrac{1}{2} \leq x \leq +\tfrac{1}{2},$$
$$-\infty \leq y \leq +\infty.$$

Construisons sur le segment $[-\tfrac{1}{2}, +\tfrac{1}{2}]$ l'ensemble P et définissons ensuite sur ce segment la fonction $\Phi(x)$.

Définissons sur l'ensemble M un système dynamique, dont le mouvement est exprimé par l'équation différentielle suivante:

$$dx/dt = 0, \qquad dy/dt = -[\,|\,y\,| + \Phi(x)\,].$$

Les seconds membres de ces équations différentielles sont des fonctions continues satisfaisant aux conditions de Lipschitz.

Les lignes sur lesquelles sont situées les trajectoires des points de l'ensemble M sont les droites:

$$x = \text{const.}$$

Les trajectoires des mouvements sont les ensembles suivants: tous les points appartenant à l'ensemble P construit sur l'axe x sont des points de repos,

vu que dans ces points les seconds membres des équations différentielles s'annulent. Les droites qui n'ont pas de parties communes avec l'ensemble P sont des trajectoires de mouvements instables au sens de Lagrange dans les deux sens. L'ensemble de points situés sur ces droites est un ensemble dense partout dans M. Les droites qui ont des parties communes avec l'ensemble P sont formées par les trajectoires de trois mouvements indépendants: du point de repos sur l'axe x et de deux semi-droites qui y sont contigues et qui sont les trajectoires de deux mouvements asymptotiques. Le point de repos est pour l'un de ces mouvements le point limite alpha, et pour l'autre—le point limite oméga du mouvement.

Désignons par M^* un ensemble de points dont les coordonnées satisfont aux inégalités $-\frac{1}{2} \leq x \leq +\frac{1}{2}$, $0 \leq y \leq 1$ et qui de plus sont situés sur des droites ayant des parties communes avec l'ensemble P. Les ponits de l'ensemble M^*, pour lesquels $y = 0$ sont des points de repos, et les points pour lesquels $y > 0$ se rapprochent asymptotiquement des points de repos pour $t \to +\infty$.

L'ensemble M^* est mesurable: $\operatorname{mes} M^* = \frac{1}{2}$. Il découle du caractère des mouvements des points de l'ensemble M^* que

$$\lim_{t \to +\infty} \operatorname{mes} f(M^*, t) = 0.$$

Ceci veut dire que l'invariant intégral de H. Poincaré ne peut être introduit dans le système dynamique éxaminé. Mais dans ce système a lieu l'incompressibilité des domaines, vu qu'il se trouvera dans chacun de ces domaines un point qui quitte ce domaine pour toujours.

MOSCOU, U. S. S. R.

THE STRUCTURE OF MONOTONE FUNCTIONS.*

By Philip Hartman and Richard Kershner.

Introduction. The present paper deals with continuous, monotone increasing functions

$$y = f(x), \text{ where } 0 \leqq x \leqq 1 \text{ and } 0 \leqq y \leqq 1,$$

and attempts a classification of the absolutely continuous or singular behavior of these functions, these notions being meant in the sense of Lebesgue. It is clear that this class of functions is in one-to-one correspondence with the class of monotone mappings of a dense enumerable subset of the interval $0 \leqq x \leqq 1$ into a dense enumerable set of the interval $0 \leqq y \leqq 1$. Hence it is natural to attempt to describe the Lebesgue properties (absolute continuity, pure "singularity," or mixed behavior) of these functions $f(x)$ directly in terms of the asymptotic or qualitative properties of the two dense sequences of numbers which are mapped on to each other by $y = f(x)$. This approach was suggested to us by Professor Wintner.

One can fix one of the two sequences arbitrarily without loss of generality. The simplest choice seems to be the dense set on the x-interval consisting of all points which have a finite dyadic development. Thus the Lebesgue properties of $f(x)$ will be explicitly described in terms of the sequence of numbers $\{y_j\}$ consisting of all function values $f(k/2^n)$, where $k = 0, 1, \cdots, 2^n$ and $n = 0, 1, \cdots$. For instance, there will be obtained (Section 1) necessary and sufficient conditions for the function $y = f(x)$ to be absolutely continuous and corresponding conditions for $y = f(x)$ to be purely singular, i. e., to have no absolutely continuous component.

Essential use will be made of the fact that, while the inverse of a strictly increasing, absolutely continuous function need not be absolutely continuous, the inverse of a strictly increasing, purely singular function is necessarily purely singular.

For a certain class of functions, which will be called the class belonging to the symmetric case, the results will be particularly complete (Sections 3 and 6). For instance, both these functions and their inverses will be shown to be "pure," i. e., either absolutely continuous or purely singular. In the treatment of the symmetric case, use is made of the "pure" theorem [1] of the

* Received April 23, 1937.

[1] Jessen and Wintner [4], Theorem XXXV.

theory of infinite convolutions and the theory of probable convergence of series of random variables. Actually, infinite products of random variables are used, but this is easily reduced to the theory of series. The latter theory, developed by Rademacher [12], Khintchine and Kolmogoroff [6] and others, is known to be equivalent to the convergence theory of infinite convolutions.[2] The theory of convolutions is made applicable by showing (Section 6) that certain of the functions of the symmetric case are infinite convolutions of the Poisson type; cf. Wintner [15]. This fact, on the one hand, makes possible a sharpening of the results of Section 3 and, on the other hand, provides an interesting theorem on infinite convolutions.

The description of the function $y = f(x)$ in terms of the sequence of numbers $k/2^n$ provides for the simple construction of particular functions of a given type, for instance, (Section 4) absolutely continuous functions which are nowhere monotone, or (Section 5) strictly increasing, purely singular functions which show that no restriction on the modulus of continuity which is weaker than a uniform Lipschitz condition of order 1 assures absolute continuity.

The theorems to be proved imply, in particular, the results and examples given by Brodén [1], Minkowski [10], Denjoy [2], Faber [3], Sierpiński [13], van Kampen and Wintner [5]. The products $\Pi(1 \pm \epsilon_n)$ occurring in the sequel have been considered by Brodén [1] and their connection with the question of singularity has been recognized by Faber [3] and more particularly by Rademacher [11].

1. The general case. Let $y = f(x)$ be a continuous, monotone increasing function defined on the interval $[0, 1]$ such that $f(0) = 0$ and $f(1) = 1$. Let $y_n{}^k$ for $k = 0, 1, \cdots, 2^n$ denote the image of the point $x = k/2^n$ so that

$$(1) \qquad f(k/2^n) = y_n{}^k; \qquad (k = 0, 1, \cdots, 2^n;\ n = 0, 1, \cdots).$$

Then the set $\{y_n{}^k\}$ is a dense set R of points on the interval $[0, 1]$, and, since $f(x)$ is monotone increasing,

$$(2) \qquad y_n{}^k \overset{<}{\underset{>}{=}} y_m{}^j \text{ according as } k/2^n \overset{<}{\underset{>}{=}} j/2^m.$$

The numbers $y_n{}^k$, where $k = 0, 1, \cdots, 2^n$, will be called the n-th section R_n of R.

Denote by $\gamma_n{}^0, \gamma_n{}^1, \cdots, \gamma_n{}^{2^n-1}$ the 2^n open intervals

$$(3) \qquad \gamma_n{}^k\colon\ y_n{}^k < y < y_n{}^{k+1}; \qquad (k = 0, 1, \cdots, 2^n - 1),$$

into which the y-interval $[0, 1]$ is divided by the points of R_n. Each interval $\gamma_n{}^k$ is subdivided into exactly two subintervals γ_{n+1}^{2k} and γ_{n+1}^{2k+1}. For convenience,

[2] Jessen and Wintner [4], Theorem XXXII.

one of these subintervals will be denoted by α_n^k, the other by β_n^k, the notation being fixed so that

$$(4) \qquad \alpha_n^k \geqq \beta_n^k \text{ and } \alpha_n^k + \beta_n^k = \gamma_n^k;\ k = 0, 1, \cdots, 2^n - 1,$$

where the same letter is used to denote either the interval or its length. Each point y of $[0, 1]$ which is not a point of R_n determines, for every n, a unique interval γ_n^k such that y is in γ_n^k. The corresponding sequence of intervals will be denoted by $\gamma_n(y)$. The intervals α_n^k and β_n^k into which the interval $\gamma_n^k = \gamma_n(y)$ is subdivided by a point of R_{n+1} will be denoted by $\alpha_n(y)$ and $\beta_n(y)$. Now let

$$(5) \qquad \begin{matrix} \epsilon_n(y) = [\alpha_n(y) - \beta_n(y)]/\gamma_n(y) \text{ if } y \text{ is in } \alpha_n(y) \\ \epsilon_n(y) = [\beta_n(y) - \alpha_n(y)]/\gamma_n(y) \text{ if } y \text{ is in } \beta_n(y) \end{matrix} \qquad (n = 0, 1, \cdots).$$

The symbol $\epsilon_n(y)$ will remain undefined if y is a point of R_{n+1}. Thus $\epsilon_n(y)$ is a step function defined over the complement of R_{n+1} in $[0, 1]$; also $-1 < \epsilon_n(y) < +1$.

THEOREM I. *The continuous, strictly increasing function $y = f(x)$ is purely singular, mixed, or absolutely continuous according as the set of points y for which the infinite product*

$$(6) \qquad \prod_{n=0}^{\infty} [1 + \epsilon_n(y)]$$

diverges,[3] *is of measure one, of measure between one and zero, or of measure zero.*

The proof of Theorem I depends on the following

LEMMA I. *A continuous function of bounded variation, $y = \phi(x)$, is purely singular, mixed, or absolutely continuous according as the variation on the set of points at which the derivative $\phi'(x)$ is not finite is equal to the total variation, less than the total variation but positive, or zero.*

This lemma is implicitly contained in the work of Lebesgue [9].

Proof of Theorem I. In order to apply Lemma I, notice that the variation over the set of points at which $f'(x)$ is not finite is equal to the measure of the set of points y for which $df^{-1}(y)/dy$ is zero, where $x = f^{-1}(y)$, $0 \leqq y \leqq 1$, is the continuous, monotone, inverse function of $y = f(x)$, $0 \leqq x \leqq 1$. Put

$$(7) \qquad h_n(y) = (\tfrac{1}{2})^n/\gamma_n(y), \qquad (n = 0, 1, \cdots),$$

[3] Divergence of an infinite product is understood to include the case of divergence to zero.

so that, by (1), (3), and the definition of $\gamma_n(y)$,

(8) $h_n(y) = [f^{-1}(y_n^{k+1}) - f^{-1}(y_n^k)]/(y_n^{k+1} - y_n^k)$, where $y_n^k < y < y_n^{k+1}$.

It is clear from (8) that $h_n(y) \to df^{-1}(y)/dy, n \to \infty$, if $df^{-1}(y)/dy$ exists. Since $f^{-1}(y)$ is monotone, its derivative exists almost everywhere on the interval $0 \leqq y \leqq 1$. Thus

$$h_n(y) \, (\to) \, df^{-1}(y)/dy, \, n \to \infty, \tag{9}$$

where the parenthesis enclosing a relation sign means that the indicated relation holds almost everywhere.

Now suppose that $\gamma_{n+1}(y) = \alpha_n(y)$; then, by (5),

$$\alpha_n(y) - \beta_n(y) = \epsilon_n(y)\gamma_n(y).$$

Hence, from (4),

$$\alpha_n(y) = \tfrac{1}{2}\gamma_n(y)[1 + \epsilon_n(y)].$$

Similarly, if $\gamma_{n+1}(y) = \beta_n(y)$, then

$$\beta_n(y) = \tfrac{1}{2}\gamma_n(y)[1 + \epsilon_n(y)].$$

Thus, in any case,

$$\gamma_{n+1}(y) = \tfrac{1}{2}\gamma_n(y)[1 + \epsilon_n(y)],$$

and so, since $\gamma_0(y) \equiv 1$,

$$\gamma_{n+1}(y) = (\tfrac{1}{2})^{n+1}\prod_{k=0}^{n}[1 + \epsilon_k(y)]. \tag{10}$$

From (7), (9) and (10),

$$1/\prod_{k=0}^{n}[1 + \epsilon_k(y)] \, (\to) \, df^{-1}(y)/dy, \qquad n \to \infty,$$

that is,

$$1/\prod_{k=0}^{\infty}[1 + \epsilon_k(y)] \, (=) \, df^{-1}(y)/dy. \tag{11}$$

Theorem I now follows from (11) and Lemma I in virtue of the remark made at the beginning of the proof.

2. A criterion for pure singularity. A weakened form of the criterion for pure singularity furnished by Theorem I may be given which lends itself readily to the explicit calculation of particular examples.

THEOREM II. *If* $\liminf_{n \to \infty} \beta_n(y)/\alpha_n(y) \, (<) \, 1$, *then* $y = f(x)$ *is purely singular.*

Proof. Since, by (5),

$$|\epsilon_n(y)| = \left| \frac{\alpha_n(y) - \beta_n(y)}{\alpha_n(y) + \beta_n(y)} \right| > \tfrac{1}{2} \left| 1 - \frac{\beta_n(y)}{\alpha_n(y)} \right| ,$$

it is clear that $\liminf_{n\to\infty} [\beta_n(y)/\alpha_n(y)] (<) 1$ implies $\limsup_{n\to\infty} |\epsilon_n(y)| (>) 0$, which in turn implies the divergence of the infinite product (6) almost everywhere on the interval $0 \leqq y \leqq 1$. Hence Theorem II follows from Theorem I.

A continuous, monotone increasing function considered by Minkowski [10] in connection with Lagrange's theorem on the continued fractions of quadratic irrationalities was proved to be purely singular by Denjoy [2]. Denjoy's proof depended on the metric properties of continued fractions. A more direct treatment can be obtained from Theorem II as follows:

For convenience, we shall denote by P_n a point of R_n which is not a point of R_{n-1}. The Minkowski function is determined by the following set R: Let R_0 consist of the points $0/1$ and $1/1$; suppose that R_n has been defined and consists of rational numbers, let p/q, p'/q' be two successive points of R_n, then $(p+p')/(q+q')$ is the point P_{n+1} between p/q and p'/q'. Note that if $p/q < p'/q'$ are two successive points of R_n, then $p'q - pq' = 1$, so that the $\gamma_n{}^k$ determined by these two points is of length $1/(qq')$. Thus the ratio $\beta_n{}^k/\alpha_n{}^k$ is q/q' or q'/q according as $q < q'$ or $q' < q$. In particular, the interval γ^{4k}_{n+2} determined by R_{n+2} whose endpoints are p/q and $(2p+p')/(2q+q')$ is divided by a point P_{n+3} in such a way that

$$\beta^{4k}_{n+2}/\alpha^{4k}_{n+2} = q/(2q+q') < \tfrac{1}{2}.$$

Hence if $\beta_n{}^k/\alpha_n{}^k \geqq \frac{1}{2}$, then neither endpoint of $\gamma_n{}^k$ is a point of R_{n-2} and so, one is a point P_{n-1} and the other a point P_n.

Now suppose that $\liminf_{n\to\infty} [\beta_n(y)/\alpha_n(y)]$ is 1 on a set of positive measure. Then, since $\beta_n(y)/\alpha_n(y) \leqq 1$, it follows that $\beta_n(y)/\alpha_n(y) \to 1$ on a set of positive measure, and so there exists a y-set T of positive measure and a sufficiently large integer N such that $\beta_n(y)/\alpha_n(y) > \frac{1}{2}$, if $n \geqq N$ and y on T. Consider the points of T which lie in a particular $\gamma_N{}^k$. Then this interval $\gamma_N{}^k$ is divided into four intervals by a point P_{N+1} and two points P_{N+2}. The above considerations show that the points of T which are in $\gamma_N{}^k$ must lie in the two intervals determined by the point P_{N+1} and a point P_{N+2}. Each of these intervals is divided into two intervals by two points P_{N+3} and again the points of T which are in $\gamma_N{}^k$ must lie in the two intervals whose endpoints are one of these P_{N+3} and the corresponding P_{N+2}. Continuing in this manner, it is seen that there can be at most two points of T in any $\gamma_N{}^k$. This contradicts the fact that T is a set of positive measure.

3. The symmetric case. The set R will be called symmetric if the ratios $\alpha_n{}^k/\beta_n{}^k$ are independent of k, i. e., if $|\epsilon_n(y)| = \epsilon_n$ is independent of y. If the larger interval $\alpha_n{}^k$ is always nearer to the point $y = 0$ than the smaller

interval $\beta_n{}^k$, the set R will be called lower symmetric. In the symmetric case, formula (5) reduces to

$$\begin{aligned} \epsilon_n(y) &= +\epsilon_n \text{ if } y \text{ is in } \alpha_n(y) \\ \epsilon_n(y) &= -\epsilon_n \text{ if } y \text{ is in } \beta_n(y) \end{aligned} \qquad (n = 0, 1, \cdots), \tag{12}$$

where $0 \leqq \epsilon_n < 1$.

The case where ϵ_n is also independent of n has been treated by Faber [3], who has shown that in this case dy/dx can take only the two values 0 and $+\infty$. These simple functions seem to be the earliest examples of nowhere constant purely singular functions.

Next there will be proved

THEOREM III. *If R is symmetric, then $x = f^{-1}(y)$ is purely singular or absolutely continuous according as $\sum_{n=0}^{\infty} \epsilon_n{}^2$ is divergent or convergent.*

Since the inverse of a purely singular function is purely singular, Theorem III implies that $y = f(x)$ is purely singular if and only if the series $\sum_{n=0}^{\infty} \epsilon_n{}^2$ is divergent. On the other hand, the inverse of an absolutely continuous function is not necessarily absolutely continuous, i. e., it may be mixed. However, if R is symmetric, then $y = f(x)$ is absolutely continuous whenever $x = f^{-1}(y)$ is absolutely continuous. Thus the statement of Theorem III remains true if the function $x = f^{-1}(y)$ is replaced by the function $y = f(x)$. The proof of this fact depends on the theory of infinite convolutions and will be deferred to Section 6.

In order to prove Theorem III several lemmas will be needed.

LEMMA II. *If $\{\epsilon_n\}$ is an arbitrary fixed sequence for which $0 \leqq \epsilon_n < 1$, then the infinite product*

$$\prod_{n=0}^{\infty} (1 \pm \epsilon_n) \tag{13}$$

is convergent almost everywhere or divergent to zero almost everywhere according as $\sum_{n=0}^{\infty} \epsilon_n{}^2$ is convergent or divergent.

The "almost everywhere" of the lemma is meant in the usual sense (cf., e. g., Steinhaus [14]). A (1, 1)-correspondence is defined between the set of sequences $\{\pm, \pm, \cdots\}$ and the points of the interval $[0, 1]$, with the exception of an enumerable set on the latter interval. The sign $+$ is made to correspond to 0, the sign $-$ to 1. Then any sequence of $+$ and $-$ signs corresponds to the dyadic development of a point of the interval $[0, 1]$. Conversely, any point of the interval $[0, 1]$ (up to an enumerable set) has a unique

dyadic development which corresponds to a sequence of $+$ and $-$ signs. A set of sequences $\{\pm, \pm, \cdots\}$ is said to be measurable if the corresponding set of points in the interval $[0, 1]$ is measurable in the sense of Lebesgue; also the measure of a measurable set of sequences $\{\pm, \pm, \cdots\}$ is defined as the measure of its image set. Then the "almost everywhere," in the statement of Lemma II, means "for all sequences $\{\pm, \pm, \cdots\}$ with the possible exception of a set of sequences which has measure zero." [4]

Proof of Lemma II. The fact that the convergence of $\sum_{n=0}^{\infty} \epsilon_n{}^2$ implies the convergence of the product (13) almost everywhere is a well-known result of Rademacher [12]. However the entire lemma may be obtained with the help of a theorem of Khintchine and Kolmogoroff [6]. Clearly, the product (13) converges if and only if the corresponding sum $\sum_{n=0}^{\infty} \log(1 \pm \epsilon_n)$ converges. Now,

[4] It may be mentioned that the fact that the product (13), for a fixed sequence $\{\epsilon_n\}$, is either convergent almost everywhere or divergent almost everywhere is a consequence of the famous "0 or 1" principle (Cf. Kolmogoroff [8]). An elementary proof of this principle for the case under consideration will be given here. This proof may be extended to a more general class of problems.

Let $\{\delta_1, \delta_2, \cdots, \delta_n, \delta_{n+1}, \cdots\}$, where δ_n is a $+$ sign or a $-$ sign, be a particular sequence for which (13) converges. Then (13) also converges for any of the sequences $\{\pm, \pm, \cdots, \pm, \delta_{n+1}, \delta_{n+2}, \cdots\}$. Now let S be the set of points of the interval $[0, 1]$ which is the image of the set of all sequences $\{\pm, \pm, \cdots\}$ for which (13) converges. The set S is clearly measurable. Let $\psi(x)$, $0 \leqq x < 1$, denote the characteristic function of S, i.e., $\psi(x)$ is 1 or 0 according as x is or is not a point of S. Let the domain of definition of $\psi(x)$ be extended to the entire x-axis by the formula $\psi(x+1) = \psi(x)$. The above property of the image of S means that $\psi(x + (\frac{1}{2})^n) = \psi(x)$ for all x and for all integers n. Thus, $\psi(x)$ is a periodic, measurable function with arbitrarily small periods. Hence, $\psi(x)$ is constant almost everywhere. Thus, S is either a zero set or a set of full measure.

Added September 28, 1937. Using the method of Kac [*Commentarii Mathematici Helvetici*, vol. 9 (1936-37), pp. 170-171], it is possible to give a simple proof of the Burstin theorem [*Monatshefte für Mathematik und Physik*, vol. 26 (1915), p. 234] employed above. The theorem states that if $g(x)$ is a measurable function having arbitrarily small periods, then $g(x)$ is a constant almost everywhere. Let $\{\tau_n\}$ be a sequence of periods of $g(x)$ such that $\tau_n \to 0$, $n \to \infty$. The expression

$$M = \tau_n{}^{-1} \int_x^{x+\tau_n} e^{ig(t)} dt$$

is independent of both x and τ_n, since it is equal to

$$M(e^{ig(t)}) = \lim_{T\to\infty} T^{-1} \int_0^T e^{ig(t)} dt.$$

Now,

$$M = \tau_n{}^{-1} \int_x^{x+\tau_n} e^{ig(t)} dt \ (\to) \ e^{ig(x)}, \ n \to \infty;$$

so that, $g(x)$ $(\equiv)$ const. (mod 2π). Application of this argument to $cg(x)$, where c is arbitrary, yields $g(x)$ $(\equiv)$ const. (mod $2\pi/c$). Hence $g(x)$ $(=)$ const.

according to Khintchine and Kolmogoro₃, the sum $\sum_{n=0}^{\infty} \log(1 \pm \epsilon_n)$ converges almost everywhere if the mathematical expectancies

$$a_n = \tfrac{1}{2} \log (1 - \epsilon_n{}^2)$$

and the mean square deviations

$$b_n = \tfrac{1}{4} \left[\log \frac{1+\epsilon_n}{1-\epsilon_n} \right]^2$$

are such that $\sum_{n=0}^{\infty} a_n$, $\sum_{n=0}^{\infty} b_n$ are convergent series; cf. also Jessen and Wintner [4], Theorems V and XXXII. These conditions are clearly satisfied if and only if the series $\sum_{n=0}^{\infty} \epsilon_n{}^2$ is convergent. Now it may be supposed that $|\log(1 \pm \epsilon_n)| < 1$, at least if $n > N$, since otherwise the sum $\sum_{n=N}^{\infty} \log(1 \pm \epsilon_n)$, and hence the product (13), is divergent everywhere. Under this assumption, a result of Kolmogoroff [7] implies that $\sum_{n=N}^{\infty} \log(1 \pm \epsilon_n)$ is divergent almost everywhere unless $\sum_{n=0}^{\infty} a_n$, $\sum_{n=0}^{\infty} b_n$ are convergent series, i. e., unless $\sum_{n=0}^{\infty} \epsilon_n{}^2$ is convergent. That the product (13) is divergent to zero almost everywhere if $\sum_{n=0}^{\infty} \epsilon_n{}^2 = +\infty$, may be seen from (15) below. This completes the proof of Lemma II.

LEMMA III. *If R is lower symmetric, then $x = f^{-1}(y)$ is purely singular or absolutely continuous according as $\sum_{n=0}^{\infty} \epsilon_n{}^2$ is divergent or convergent.*

Proof. Suppose y is a point of the interval $[0, 1]$ but not a point of the set R. Then y determines a unique product

$$\prod_{n=0}^{\infty} [1 + \epsilon_n(y)] = \prod_{n=0}^{\infty} [1 + \epsilon_n(f(x))] = \prod_{n=0}^{\infty} (1 \pm \epsilon_n). \tag{14}$$

Thus y determines uniquely a particular sequence $\{\pm, \pm, \cdots\}$, by (12). On the other hand, this sequence $\{\pm, \pm, \cdots\}$ corresponds to the dyadic development of a point x of the interval $[0, 1]$. The fact that R is lower symmetric clearly implies that this x satisfies $y = f(x)$. Thus the convergence or divergence of (13) "almost everywhere" in the sense of Lemma II means the convergence or divergence of $\prod_{n=0}^{\infty} [1 + \epsilon_n(f(x))]$ "almost everywhere" in

the usual sense, i. e., for all x in $[0, 1]$ with the possible exception of a zero set. Now, by (11) and (14),

$$\prod_{n=0}^{\infty} [1 + \epsilon_n(f(x))] (=) f'(x). \tag{15}$$

In virtue of (15), the proof of Lemma III is completed by an application of Lemma I and Lemma II.

Proof of Theorem III. Let x be a point of the interval $[0, 1]$ which does not possess a finite dyadic development. This x determines a unique y by the condition $y = f(x)$. Also, $y = f(x)$ determines a unique sequence $\{\pm, \pm, \cdots\}$, by (12). This sequence corresponds to the dyadic development of a point $\bar{x}$ of the interval $[0, 1]$. Thus, there is determined, up to an enumerable set, a measure preserving $(1, 1)$-transformation of the interval $[0, 1]$ into itself by the formula $x \to \bar{x}$. Now the symmetric set R determines a sequence of numbers $\{\epsilon_n\}$. Let $\bar{R}$ be the lower symmetric set which is uniquely determined by this sequence $\{\epsilon_n\}$ and let $y = \bar{f}(x)$, $0 \leqq x \leqq 1$, denote the corresponding function. Then,

$$\prod_{n=0}^{\infty} [1 + \epsilon_n(f(x))] (=) \prod_{n=0}^{\infty} [1 + \epsilon_n(\bar{f}(\bar{x}))]. \tag{16}$$

On the other hand, by (11),

$$f'(x) (=) \prod_{n=0}^{\infty} [1 + \epsilon_n(f(x))] \text{ and } \bar{f}'(x) (=) \prod_{n=0}^{\infty} [1 + \epsilon_n(\bar{f}(\bar{x}))]. \tag{17}$$

Thus, by (16) and (17),

$$f'(x) (=) \bar{f}'(\bar{x}).$$

So, by Lemma I and the fact that the correspondence $x \to \bar{x}$ is measure preserving, it is seen that the function $x = f^{-1}(y)$ is purely singular or absolutely continuous according as the function $\bar{x} = \bar{f}^{-1}(y)$ is purely singular or absolutely continuous. Theorem III now follows from Lemma III.

4. Nowhere monotone, absolutely continuous functions. Theorem III makes possible the construction, in a simple way, of functions which are absolutely continuous but are not monotone in any interval. Let R be a symmetric set associated with a sequence $\{\epsilon_n\}$, $0 \leqq \epsilon_n < 1$, for which $\sum_{n=0}^{\infty} \epsilon_n^2 < +\infty$ but $\sum_{n=0}^{\infty} \epsilon_n = +\infty$. By Theorem III, the function $x = f^{-1}(y)$ is absolutely continuous, where $y = f(x)$ is the function defined by R. Let γ_m^k be an arbitrary interval determined by R_m. This γ_m^k determines a set $(\delta_0, \delta_1, \cdots, \delta_m)$

of + or — signs, which are the first $(m+1)$ + or — signs determined by any y in $\gamma_m{}^k$. Now, since $\sum_{n=0}^{\infty} \epsilon_n = +\infty$, it is clear that there exist at least two sequences $\{\delta_0, \delta_1, \cdots, \delta_m, \pm, \pm, \cdots\}$ such that the product (13) diverges to $+\infty$ if the $\pm$ signs in the product are chosen to be those of one of these sequences and diverges to 0 if the $\pm$ signs are those of the other sequence. On the other hand, the reciprocal value of $\prod_{n=0}^{\infty} [1 + \epsilon_n(y)]$ is a limit of difference quotients for the function $x = f^{-1}(y)$ at the point y. Thus in the interval $\gamma_m{}^k$ there exist at least two points such that the function $x = f^{-1}(y)$ has a derivative number equal to $+\infty$ at one of these points and at the other a derivative number equal to 0. Obviously, the function $x = f^{-1}(y) - y$ is absolutely continuous but not monotone in any y-interval.

In virtue of the remark following the statement of Theorem III, it may be seen that the function $y = f(x) - x$ is also absolutely continuous and nowhere monotone.

5. The modulus of continuity of purely singular functions. Let $y = f(x)$ be a function defined on the interval $[0, 1]$; let

$$\omega(\delta) = \max | f(x_1) - f(x_2) |$$

for all x_1, x_2 belonging to $[0, 1]$ and such that $| x_1 - x_2 | \leqq \delta$. Then the function $\omega(\delta)$ is called the modulus of continuity of $f(x)$. The fact that $y = f(x)$ satisfies the Lipschitz condition of order α may be expressed by the formula $\omega(\delta) = O(\delta^\alpha)$. If $f(x)$ satisfies the Lipschitz condition of order one, $\omega(\delta) = O(\delta)$, then $f(x)$ is necessarily absolutely continuous. The object of the present section is to show that in a certain sense this statement cannot be improved.

For every monotone decreasing function $\phi(\delta)$ which approaches $+\infty$ as $\delta \to +0$, there exists a continuous, strictly increasing, purely singular function, $y = f(x)$, for which $\omega(\delta) = O(\delta\phi(\delta))$.

Proof. The function $y = f(x)$ will be chosen to be of the type considered in Section 3, i. e., it will be defined by the construction of a symmetric set R. In fact, every ϵ_n will be chosen to be either 0 or a fixed constant c, $0 < c < 1$. If ϵ_n is chosen in this manner, the set R will clearly be dense. Furthermore, by Theorem III, $f(x)$ will be purely singular if $\epsilon_n = c$ for infinitely many values of n.

Now it is clear that

$$\omega(2^{-n}) \leqq 2 \max_k \gamma_n{}^k = 2 \cdot 2^{-n} \prod_{j=0}^{n-1} (1 + \epsilon_j).$$

Thus we obtain

$$(18) \qquad \omega(2^{-n})/2^{-n}\phi(2^{-n}) \leqq 2 \prod_{j=0}^{n-1} (1 + \epsilon_j)/\phi(2^{-n}).$$

Now, since $\phi(2^{-n})$ increases indefinitely with n, there exists a monotone sequence of positive integers $\{k_n\}$ such that

$$(19) \qquad \phi(2^{-k_n}) > (1 + c)^n, \qquad (n = 0, 1, \cdots).$$

Now put $\epsilon_j = c$ or 0 according as j does or does not belong to the sequence $\{k_n\}$. Then

$$(20) \qquad \prod_{j=0}^{n-1} (1 + \epsilon_j) = (1 + c)^i, \text{ where } k_i \leqq n - 1 < k_{i+1}.$$

Thus it follows from (18), (19), (20) and the monotony of $\phi(\delta)$ that

$$(21) \qquad \omega(2^{-n})/2^{-n}\phi(2^{-n}) \leqq 2(1 + c)^i/\phi(2^{-k_i}) \leqq 2.$$

Now let $2^{-(n-1)} < \delta \leqq 2^{-n}$, then $\omega(\delta) \leqq \omega(2^{-n})$. Also

$$(22) \qquad \omega(\delta)/\delta \leqq [\omega(2^{-n})/2^{-n}][2^{-n}/\delta] \leqq 2[\omega(2^{-n})/2^{-n}].$$

On the other hand, $\phi(\delta) > \phi(2^{-n})$, so that, by (21) and (22),

$$\omega(\delta)/\delta\phi(\delta) \leqq 2\omega(2^{-n})/2^{-n}\phi(2^{-n}) \leqq 4.$$

This completes the proof of the italicized statement above.

6. Certain Poisson convolutions. Let $\sigma_n = \sigma_n(x)$ denote the distribution function

$$(23) \qquad \begin{aligned} \sigma_n(x) &= 0, & -\infty < x \leqq 0; \\ \sigma_n(x) &= \tfrac{1}{2}(1 + \epsilon_n), & 0 < x \leqq (\tfrac{1}{2})^{n+1}; \\ \sigma_n(x) &= 1, & (\tfrac{1}{2})^{n+1} < x < +\infty; \end{aligned}$$

$0 \leqq \epsilon_n < 1$; $n = 0, 1, \cdots$. Thus $\sigma_n(x)$ is the step function with the jump $\frac{1}{2}(1 + \epsilon_n)$ at $x = 0$ and the jump $\frac{1}{2}(1 - \epsilon_n)$ at $x = (\frac{1}{2})^{n+1}$. Clearly, the convolution

$$\tau_1(x) = \sigma_0 * \sigma_1 = \int_{-\infty}^{+\infty} \sigma_0(x - \xi)\, d\sigma_1(\xi)$$

is the step function with the jumps

$$\begin{aligned} &(\tfrac{1}{2})^2(1 + \epsilon_0)(1 + \epsilon_1) \text{ at } x = 0, \\ &(\tfrac{1}{2})^2(1 + \epsilon_0)(1 - \epsilon_1) \text{ at } x = (\tfrac{1}{2})^2, \\ &(\tfrac{1}{2})^2(1 - \epsilon_0)(1 + \epsilon_1) \text{ at } x = 2(\tfrac{1}{2})^2, \text{ and} \\ &(\tfrac{1}{2})^2(1 - \epsilon_0)(1 - \epsilon_1) \text{ at } x = 3(\tfrac{1}{2})^2. \end{aligned}$$

Similarly, the convolution

$$(24) \qquad \tau_n(x) = \sigma_0 * \sigma_1 * \cdots * \sigma_n = \tau_{n-1} * \sigma_n$$

has jumps $(\frac{1}{2})^{n+1} \prod_{j=0}^{n} (1 \pm \epsilon_j)$ at the points $k/2^{n+1}$, $k = 0, 1, \cdots, 2^{n+1} - 1$. The arrangement of $+$ and $-$ signs in the product which represents the jump of τ_n at the point $k/2^{n+1}$ is the same as the arrangement of zeros and ones in the finite dyadic development of the point $k/2^{n+1}$. Let γ^k_{n+1} denote the jump

$$(25) \qquad \gamma^k_{n+1} = (\tfrac{1}{2})^{n+1} \prod_{j=0}^{n} (1 \pm \epsilon_j), \qquad (k = 0, 1, \cdots, 2^{n+1} - 1)$$

of τ_n at the point $x = k/2^{n+1}$. Let

$$(26) \qquad y^k_{n+1} = \sum_{j<k} \gamma^j_{n+1}, \qquad (k = 0, 1, \cdots, 2^{n+1}),$$

so that

$$(27) \qquad \tau_n(k/2^{n+1}) = y^k_{n+1}.$$

Now, by (25) and the remark concerning the arrangement of $+$ signs and $-$ signs in (25), it is clear that

$$\gamma^{2k}_{n+2} = \tfrac{1}{2}\gamma^k_{n+1}(1 + \epsilon_{n+2}) \quad \text{and} \quad \gamma^{2k+1}_{n+2} = \tfrac{1}{2}\gamma^k_{n+1}(1 - \epsilon_{n+2}),$$

so that

$$\gamma^{2k}_{n+2} + \gamma^{2k+1}_{n+2} = \gamma^k_{n+1}.$$

Thus, by (26),

$$(28) \qquad y^k_{n+1} = y^{2k}_{n+2}.$$

From (27) and (28) it is seen that

$$(29) \qquad \tau_m(k/2^{n+1}) = y^k_{n+1}, \qquad (m \geqq n).$$

Now let $\tau(x)$ denote the infinite convolution

$$(30) \qquad \tau(x) = \sigma_0 * \sigma_1 * \cdots = \lim_{n \to \infty} \tau_n(x).$$

The conditions for the convergence of this infinite convolution are satisfied in the case under consideration for any choice of $\{\epsilon_n\}$, $0 \leqq \epsilon_n < 1$; cf. Jessen and Wintner [4], Theorem V. It will be supposed that

$$(31) \qquad (\tfrac{1}{2})^{n+1} \prod_{j=0}^{n} (1 + \epsilon_j) \to 0, \qquad (n \to \infty).$$

This assures the continuity of $\tau(x)$. Now, comparing (29) with (30),

$$(32) \qquad \tau(k/2^{n+1}) = y^k_{n+1}.$$

Thus, by (1), the set of points y^k_{n+1} is the section R_{n+1} of the set R associated with $\tau(x)$. Furthermore, by (26), the γ^k_{n+1} defined by (25) are exactly the intervals γ^k_{n+1} determined by R_{n+1} (cf. (3)). Comparing (10), (12), (25) and the definition of a lower symmetric set R, it is seen that the set R associated with the function (30) is lower symmetric. Conversely, any continuous, strictly increasing function associated with a lower symmetric set R is the infinite convolution of Poisson distribution functions (23).

It has been shown by Jessen and Wintner ([4], Theorem XXXV) that a continuous infinite convolution of step functions is either purely singular or absolutely continuous. Thus a function $y = \tau(x)$ for which the corresponding set R is lower symmetric is either purely singular or absolutely continuous. On the other hand, it has been shown in Section 3, Lemma III, that $y = \tau(x)$ is purely singular if and only if $\sum_{n=0}^{\infty} \epsilon_n^2 = +\infty$. It follows that $y = \tau(x)$ is absolutely continuous if and only if $\sum_{n=0}^{\infty} \epsilon_n^2 < +\infty$.

By the use of a measure preserving transformation on the y-axis similar to that introduced in the proof of Theorem III, this statement can be extended to apply to the general symmetric case.

THEOREM IV. *If R is symmetric, then $y = f(x)$, together with $x = f^{-1}(y)$, is purely singular or absolutely continuous according as $\sum_{n=0}^{\infty} \epsilon_n^2$ is divergent or convergent.*

It should be noted that, in the definition (23) of $\sigma_n(x)$, it was not essential to restrict ϵ_n to be positive. In fact, let $\rho_n(x)$ be any distribution function of the form

$$\rho_n(x) = 0, \qquad -\infty < x \leqq 0;$$
$$\rho_n(x) = q_n, \qquad 0 < x \leqq (\tfrac{1}{2})^{n+1}; \tag{33}$$
$$\rho_n(x) = 1, \qquad (\tfrac{1}{2})^{n+1} < x < +\infty;$$

$0 < q_n < 1$; $n = 0, 1, \cdots$. Then, if the infinite convolution $\rho_0 * \rho_1 * \cdots$ is continuous, it is of the type considered in Theorem IV. Thus

THEOREM V. *The infinite Poisson convolution $\rho_0 * \rho_1 * \cdots$ of distribution functions (33) is continuous if and only if $\prod_{j=0}^{n} \max(q_j, 1 - q_j) \to 0$, $n \to +\infty$. If continuous, it is purely singular or absolutely continuous, along with its inverse, according as $\sum_{j=0}^{\infty} (q_j - \tfrac{1}{2})^2$ is divergent or convergent.*

THE JOHNS HOPKINS UNIVERSITY,
THE UNIVERSITY OF WISCONSIN.

BIBLIOGRAPHY.

[1] T. Brodén, "Beiträge zur Theorie der stetigen Functionen einer reellen Veränderlichen," *Journal für die reine und angewandte Mathematik*, vol. 118 (1897), pp. 1-60.

[2] A. Denjoy, "Sur une fonction de Minkowski," *Comptes Rendus* (Paris), vol. 198 (1934), pp. 44-47.

———, "Sur quelques points de la théorie des fonctions," *Comptes Rendus* (Paris), vol. 194 (1932), pp. 44-46.

[3] G. Faber, "über stetige Funktionen II," *Mathematische Annalen*, vol. 69 (1910), pp. 372-443; cf. pp. 395-400.

[4] B Jessen and A. Wintner, "Distribution functions and the Riemann zeta function," *Transactions of the American Mathematical Society*, vol. 38 (1935), pp. 44-88.

[5] E. R. van Kampen and A. Wintner, "On a singular monotone function," *Journal of the London Mathematical Society* (in press).

[6] A. Khintchine and A. Kolmogoroff, "Über Konvergenz von Reihen, deren Glieder durch den Zufall bestimmt werden," *Recueil Mathématique de Moscou*, vol. 32 (1925), pp. 668-677.

For another treatment and for further results, see Kolmogoroff [7].

[7] A. Kolmogoroff, "über die Summen durch den Zufall bestimmter unabhängiger Grössen," *Mathematische Annalen*, vol. 99 (1928), pp. 309-319 and vol. 102 (1930), pp. 484-488.

[8] A. Kolmogoroff, "Grundbegriffe der Wahrscheinlichkeitsrechnung," *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 2, Heft 3; in particular, see the appendix.

[9] H. Lebesgue, *Leçons sur l'intégration*, 2nd edition. In particular, see chap. IX, section 1.

[10] H. Minkowski, "Zur Geometrie der Zahlen," *Verhandlungen des III. Internationalen Mathematiker-Kongresses in Heidelberg*, 1904, pp. 164-173.

[11] H. Rademacher, "Zu dem Borelschen Satz über die asymptotische Verteilung der Ziffern in Dezimalbrüchen," *Mathematische Zeitschrift*, vol. 2 (1918), pp. 306-311; cf. also vol. 3 (1919), p. 317.

[12] H. Rademacher, "Einige Sätze über Reihen von allgemeinen Orthogonalfunktionen," *Mathematische Annalen*, vol. 87 (1922), pp. 112-138.

[13] W. Sierpiński, "Un exemple élémentaire d'une fonction croissante qui a presque partout une dérivée nulle," *Giornale de Matematiche di Battaglini*, vol. 54 (1916), pp. 314-334.

[14] H. Steinhaus, "Les probabilités dénombrables et leur rapport à la théorie de la mesure," *Fundamenta Mathematicae*, vol. 4 (1923), pp. 286-310.

[15] A. Wintner, "On convergent Poisson convolutions," *American Journal of Mathematics*, vol. 57 (1935), pp. 827-838.

PLANAR GRAPHS WHOSE HOMEOMORPHISMS CAN ALL BE EXTENDED FOR ANY MAPPING ON THE SPHERE.*

By V. W. Adkisson and Saunders MacLane.

Introduction. Certain conditions have been found under which a given homeomorphism of a given point set on a sphere can be extended to the sphere.[1]

It is the purpose of this paper to characterize the cyclicly connected planar graphs[2] G such that in every map of G on the sphere every homeomorphism of G into itself can be extended to the sphere.[3] The characterization will use only internal properties of the graph G, defined in G without reference to a map of G on any other space.

It is known that the extendability of a homeomorphism of a map of G depends on the behavior of the complementary domain boundaries of the map under the homeomorphism. Therefore the procedure for our problem is a study to determine which simple closed curves of a given graph can be boundaries of complementary domains of a map of a graph.

Definitions. The circuit[4] J of a cyclicly connected graph G is called a *bounding circuit* of G provided that for any two distinct maximal connected components, N_1 and N_2, of $G - J$ the sets $\bar{N}_1 \cdot J$ and $\bar{N}_2 \cdot J$ lie respectively on two distinct arcs AXB and AYB of J.[5]

* Presented to the Society, December 30, 1936. Received by the Editors February 6, 1937; revised April 8, 1937.

[1] H. M. Gehman, "On extending a homeomorphism between two subsets of spheres," *Bulletin of the American Mathematical Society*, vol. 42 (1936), pp. 79-81; V. W. Adkisson, "Cyclicly connected continuous curves whose complementary domain boundaries are homeomorphic," *Comptes Rendus des séances de la Société des Sciences et des Lettres de Varsovie*, vol. 23 (1930), Classe III, pp. 164-193. (Referred to in this paper as Thesis); V. W. Adkisson, "On extending a continuous $(1-1)$ correspondence of continuous curves on a sphere," *ibid.*, vol. 27 (1934).

[2] For a discussion of cyclicly connected curves in general see G. T. Whyburn, "Cyclicly connected continuous curves," *Proceedings of the National Academy of Sciences*, vol. 13 (1927), pp. 31-38. A cyclicly connected graph is also non-separable. See H. Whitney, "Non-separable and planar graphs," *Transactions of the American Mathematical Society*, vol. 34 (1932), p. 339.

[3] This paper is the outgrowth of a problem originally suggested by Professor J. R. Kline.

[4] A circuit is a simple closed curve.

[5] "Bounding circuit" is equivalent to "*boundary curve*" as defined by S. Claytor,

Note that each set $\bar{N}_i \cdot J$ consists of only a finite number of vertices of G. These vertices may be called "feet" of N_i.

A *split-circuit* of G is a bounding circuit J such that $G - J$ contains at least two components. The chief result of this paper may be stated in the following theorem.

THEOREM C. *Every homeomorphism of the cyclicly connected graph G into itself is extendable in every map of G on the sphere if, and only if, for each split circuit J and every homeomorphism σ of G such that $\sigma(G) = G$ and $\sigma(J) \neq J$, the sets $(G - J)$ and $J \cdot \sigma(J)$ each consists of two maximal connected components.*

The final theorem (Theorem D) characterizes more precisely the combinatorial structure of the graphs G satisfying the conditions of Theorem C. Specifically, such a graph either is a graph with just one map on the sphere or else is a graph with at most four essentially distinct homeomorphisms.

LEMMA 1. *Let the vertices p and q of a graph G lie on a circuit J of G, and in a map of G on a sphere S let R be one of the regions of S bounded by J. Then there is a complementary domain*[6] *boundary of G which contains both p and q and which lies in $\bar{R}$ if, and only if, there is no open arc of G in R which has ends on J and whose ends separate p and q on J.*[7]

THEOREM A. *If J is a circuit of G, then G can be mapped on a sphere so that J is a complementary domain boundary of G if, and only if, J is a bounding circuit.*[8]

This theorem can be proved by the following procedure. Take any map of G in which a certain component of $G - J$, say N, lies "outside" J.[9] One

"Topological immersion of peanian continua in a spherical surface," *Annals of Mathematics*, vol. 35 (1934), p. 809. A simple closed curve J of G is called a boundary curve of G provided that there do not exist in $G - J$ distinct components N_1 and N_2 such that (1) a point-pair of $\bar{N}_1 \cdot J$ separates a point pair of $\bar{N}_2 \cdot J$ on J, or (2) $\bar{N}_1 \cdot J = \bar{N}_2 \cdot J =$ three distinct points. Both definitions will be found useful in later proofs.

[6] A "complementary domain" will refer always to a domain complementary to some given map of G on a sphere. "Complementary domain boundary" will be abbreviated as "c. d. b."

[7] This lemma follows from a lemma by Kuratowski, "Sur les courbes gauches," *Fundamenta Mathematicae*, vol. 15 (1930), p. 274, Lemma III′; it can also be proved directly for graphs by more elementary combinatorial arguments.

[8] This is Proposition K of Claytor, *loc. cit.*, p. 828.

[9] It is sometimes convenient to refer to the two regions of a sphere bounded by a circuit J as "outside" and "inside" of J.

is able to show then, from the definition of a split-circuit, that the feet of N all lie on one c. d. b. of G "inside" and on J. Hence the given map can be altered by mapping N inside J instead of outside J. Proceeding in this manner we obtain a new map of G with all of G inside and on J.

COROLLARY 1. *If there is only one component of $G - J$ the circuit J is a c. d. b. of any map of G on a sphere.*

COROLLARY 2. *If J is a split circuit, G can be mapped so that J is not a c. d. b. of G.*

COROLLARY 3. *The c. d. b. J of a map of G will have a disconnected intersection with some other c. d. b. of the map of G if, and only if, J is a split circuit.*

Proof. The condition is sufficient. If J is a split circuit there are at least two components, N_1 and N_2, of $G - J$, while J is composed of two arcs AXB and AYB, such that $\bar{N}_1 \cdot J$ lies in AXB and $\bar{N}_2 \cdot J$ in AYB. Since G is cyclic, $\bar{N}_1 \cdot J$ is not a point, and A and B can be so chosen as to both belong to $\bar{N}_1 \cdot J$. Let R be the complementary domain of G bounded by J and set $R_1 = S - \bar{R}$. Now A and B both lie on a c. d. b. K of G with $K \subset \bar{R}_1$. For if not, by Lemma 1 there would exist a component N of $G - J$ such that two points of $\bar{N} \cdot J$ separate the points A and B on J and J would not be a split circuit. Since there is a component N_2 of $G - J$ such that $\bar{N}_2 \cdot J \subset AYB$, the arc AYB cannot be a subset of K, except possibly when $\bar{N}_2 \cdot J = A + B$. In this case a different choice of K readily gives the result. In like manner the arc AXB is not a subset of K. Therefore, since A and B are common to both K and J, $K \cdot J$ is disconnected.

The condition is necessary. Suppose G is mapped so that the c. d. b. J of G and the c. d. b. J_1 of G have a disconnected intersection F. Let F_1 and F_2 be two points in separate maximal connected subsets of this intersection. Then the removal of F_1 and F_2 disconnects G. For let the open arc $\langle F_1XF_2\rangle$ be any arc lying in the domain complementary to G bounded by J and $\langle F_1YF_2\rangle$ an arc lying in the domain complementary to G bounded by J_1. The simple closed curve $F_1XF_2YF_1$ has only the points F_1 and F_2 in common with G and disconnects G. Then F_1 and F_2 separate G into two components and neither component consists of a single arc, for otherwise F_1 and F_2 would belong to the same connected subset of F, contrary to assumption. Therefore there must exist two or more components of $G - J$, and since J is a c. d. b. of G, Theorem A shows J to be a bounding circuit. Then J is a split circuit.

THEOREM B. *In any mapping of G on a sphere S there are two complementary domain boundaries of G with a disconnected intersection if, and only if, G contains a split circuit.*

Proof. The condition is necessary from Corollary 3.

The condition is sufficient. Suppose G contains a split circuit J. Let N_1 and N_2 be two components of $G - J$ and divide J relative to N_1 and N_2 into two arcs AXB and AYB with A and B in $\bar{N}_1 \cdot J$, as in the proof of Corollary 3. Since J is a split circuit, every component M of $G - J$ satisfies either $\bar{M} \cdot J \subset AXB$ or $\bar{M} \cdot J \subset AYB$. Hence $G - (A + B)$ consists of two mutually separated sets; $H_1 = \langle AXB \rangle + N_1$ plus all components M, except N_1, with $\bar{M} \cdot J \subset AXB$, and $H_2 = \langle AYB \rangle + N_2$ plus all remaining components of $G - J$. Arcs of both sets H_1 and H_2 end on A, and so among the c. d. b.'s of G passing through A there will be at least one c. d. b. K containing arcs in both H_1 and H_2. Thus K must pass through A and B, the only points common to $\bar{H}_1$ and $\bar{H}_2$. Since A and B disconnect G into H_1 and H_2, neither one an arc, K also disconnects G, so that K is a split circuit. It is also a c. d. b. of G, so that by Corollary 3 it has a disconnected intersection with some other c. d. b., as asserted.

Proof of Theorem C. The conditions are necessary. First suppose $\sigma(J)$ does not contain any points of a component N_1 of $G - J$. Take a map of G in which J is a c. d. b. of G. As σ is extendable, $\sigma(J)$ must also be a c. d. b. of this map. We can map N_1 in the complementary domain of G bounded by J and leave $G - N_1$ fixed. But this gives a new map of G in which J is not a c. d. b. while $\sigma(J)$ remains a c. d. b. But then σ^{-1} cannot be extended to the sphere,[10] contrary to assumption. Hence $\sigma(J)$ contains points in every component of $G - J$.

Since there are at least two components N_1 and N_2 of $G - J$, $\sigma(J)$ to pass from N_1 to N_2 must contain points in J and also points not in J. Let $J \cdot \sigma(J)$ consist of k connected pieces, $g_1, g_2, \cdots, g_k$, each piece an arc or a point. Then $J - J \cdot \sigma(J)$ consists of k arcs $F_1, F_2, \cdots, F_k$, where the notation is chosen so that

$$J = F_1 + g_1 + F_2 + g_2 + \cdots + F_k + g_k \tag{1}$$

in the cyclic order shown. Because J and $\sigma(J)$ are both c. d. b.'s of G, it follows readily that the points of $g_1, \cdots, g_k$ are arranged in the same cyclic

[10] Thesis, Theorem 2. If M is a cyclicly connected continuous curve lying on a sphere S, and T a homeomorphism such that $T(M) = M$, a necessary and sufficient condition that T be extendable to S is that for every c. d. b. J of M, $T(J)$ be also a c. d. b.

order on $\sigma(J)$ as on J. Hence the arcs of $\sigma(J) - J \cdot \sigma(J)$ can be so denoted by $f_1, f_2 \cdots f_k$ that the cyclic order of arcs on $\sigma(J)$ is

$$\sigma(J) = f_1 + g_1 + f_2 + g_2 + \cdots + f_k + g_k. \tag{2}$$

The ends of f_i can then be so denoted by A_i and B_i that these points appear on J and on $\sigma(J)$ in the cyclic order $A_1B_1A_2B_2 \cdots A_kB_k$.

The circuits J and $\sigma(J)$ together divide the sphere into $k+2$ regions with the boundaries J, $\sigma(J)$, and $f_i + F_i$ $(i = 1, \cdots, k)$. As J and $\sigma(J)$ are c. d. b.'s of the map of G, the rest of G lies in the regions bounded by $f_i + F_i$. But each component of $G - J$ must contain some points of $\sigma(J)$ and hence some arc f_i. Therefore each closed region bounded by an $f_i + F_i$ contains exactly one component N_i of $G - J$, and there are k such components. Because σ carries J into $\sigma(J)$, $G - \sigma(J)$ must also have k components; namely, $N_i + F_i - (f_i + A_i + B_i)$.

Suppose now that there were three (or more) components of $G - J$. We can map N_1 in the complementary domain of G bounded by J, leaving the rest of G unaltered. This simply interchanges f_1 and F_1 in the boundaries J and $\sigma(J)$, so that there results a map of G with the c. d. b.,

$$K = f_1 + F_2 + \cdots + F_k + (J \cdot (\sigma J)), \quad L = F_1 + f_2 + \cdots + f_k + (J \cdot (\sigma J)).$$

These are the only c. d. b. containing $J \cdot \sigma(J)$, because every other c. d. b. is contained in one of $N_i + F_i$. Since σ must also be extendable in this map, it must carry K into K or L. But σ carries J into $\sigma(J)$, hence carries each F_i into some f_j, and carries K into a c. d. b. containing $\sigma(f_1)$, $k-1$ of the f_j, and $J \cdot (\sigma J)$. As this c. d. b. can be only L, and F_1 is in L, we have $\sigma(f_1) = F_1$. By the same argument, $\sigma(f_i) = F_i$ for every i.

But σ must leave fixed each end A_1 and B_1 of f_1, for otherwise $\sigma(A_1) = B_1$, $\sigma(B_1) = A_1$ which is impossible since $\sigma(A_i + B_i) = A_i + B_i$ $(i = 1, 2, \cdots, k)$ and $k \geqq 3$. This contradiction shows $\sigma A_1 = A_1$, $\sigma B_1 = B_1$.

Then also $\sigma f_1 = F_1$, $\sigma(F_1) = f_1$ and σ must carry $N_1 + F_1$ into itself. There is then another homeomorphism σ', equal to σ on N_1 and to the identity elsewhere. This σ' carries the c. d. b. J into $(J - F_1) + f_1$, which is not a c. d. b., so that σ' is not extendable in this map, contrary to hypothesis. The assumption of more than two components of $G - J$ is thus inconsistent. Therefore $G - J$ consists of two pieces N_1 and N_2, and we have already shown that $J \cdot \sigma(J)$ consists of two connected pieces, namely two arcs (or points) B_1A_2 and B_2A_1. Therefore the conditions of the theorem are necessary.

The conditions are sufficient. We need only show for each bounding

circuit J that in any map in which J is a c. d. b. of G, $\sigma(J)$ must be a c. d. b. of G. (Thesis, Theorem 2). Consider a map in which J is a c. d. b. of G. If J is not a split circuit then $\sigma(J)$ is not a split circuit and must be a c. d. b. of G. (Corollary 1, Theorem A). If $\sigma(J) = J$ the case is trivial. Hence, assume $\sigma(J) \neq J$. Denote the two components of $J \cdot \sigma(J)$ by α and β. Then J consists of four parts F_1, α, F_2, β in that cyclic order, where F_1 and F_2 are arcs, α and β may be arcs or points. The circuit $\sigma(J)$ consists of four arcs f_1, α, f_2, β in that cyclic order, where f_i and F_i have end points in common. Hence $J + \sigma(J)$ divides the sphere into four regions, r_1, r_2, r_3, r_4, with the respective boundaries: J, $\sigma(J)$, $f_1 + F_1$, and $f_2 + F_2$.

Now the arcs F_1 and F_2 of J belong to distinct components of $G - \sigma(J)$. For otherwise there would be an arc f of $G - \sigma(J)$ joining F_1 to F_2, and it readily follows that part of this arc f must lie in the region r_1, contrary to the fact that r_1 is a complementary domain of G.

Suppose now that $\sigma(J)$ is not a c. d. b. of G. Then there are arcs of G in r_2 bounded by $\sigma(J)$, so that at least one component of $G - \sigma(J)$ is contained in this region. But there are at least two other components of $G - \sigma(J)$ (those containing F_1 and F_2) not in r_2. Hence there are at least three components of $G - \sigma(J)$. But this is impossible since there are only two components of $G - J$ by hypothesis, and σ carries components of $G - J$ into components of $G - \sigma(J)$. Therefore, $\sigma(J)$ is a c. d. b. of G and the theorem is proved.

We shall now obtain other necessary and sufficient conditions that every homeomorphism of G be extendable in every map. These new conditions will give in some detail a more complete characterization of the graphs G.

LEMMA 2. *If every homeomorphism of G into itself is extendable in every map on the sphere, and if for some homeomorphism σ, some split circuit J and some component N_1 of $G - J$, $\sigma(J) = J$ and $\sigma(N_1) = N_1$, then σ has two fixed points on J.*

Proof. Because J is a split circuit we can choose an arc $F = AYB$ of J such that F contains only A and B from $\bar{N}_1 \cdot J$, while F contains $\bar{N}_2 \cdot J$ for some component N_2 of $G - J$. Map G with N_1 inside J and all other components of $G - J$ outside J. There is then inside J a domain complementary to G whose boundary K contains F. Furthermore $K = F + f$, where f is an open arc of N_1.

Suppose first that $\sigma(K) = K$. Then $\sigma(J) = J$ implies that $\sigma(f) = f$, $\sigma(F) = F$ and hence that $\sigma(A + B) = A + B$. If A and B remain fixed

we have the two desired fixed points. If A and B are interchanged by σ, there is a fixed point on[11] $F = AYB$ and also a fixed point on $J - F$.

Now assume $\sigma(K) \neq K$. The component N_2 is separated from $G - N_2$ by F and hence by $K = F + f$. Therefore K is a split circuit and N_2 a component of $G - K$. Theorem C applied to K shows that $\sigma(K)$ contains an arc of N_2. But this is impossible since $\sigma(K)$ consists of an arc of $\sigma(J) = J$ and an arc of $\sigma N_1 = N_1$, neither in N_2. Therefore $\sigma(K) = K$ and the lemma is proved.

Two homeomorphisms σ and τ are considered identical if $\sigma(G) = G$, $\tau(G) = G$ such that branch points that correspond under σ also correspond under τ, and edges that correspond under σ correspond under τ. (An edge of G means an arc of G joining two branch points, but containing no other branch points.)

THEOREM D. *Every homeomorphism of a cyclicly connected graph G into itself is extendable in every map of G on the sphere if, and only if, at least one of the following conditions holds:*

I. *Every split circuit of G is invariant under every homeomorphism σ. If there is a split circuit, then there is at most one $\sigma \neq 1$.*

II. *No split circuit is invariant under any $\sigma \neq 1$; there is only one homeomorphism $\sigma \neq 1$ and this σ is of order 2. For any split circuit J, $G - J$ consists of two connected pieces N_1 and N_2; $J \cdot \sigma(J)$ consists of just two connected pieces M_1 and M_2, and $\sigma(N_1) =$ a subset of N_1 plus an arc of J, $\sigma(N_2) =$ a subset of N_2 plus an arc of J, $\sigma(M_1) = M_2$, $\sigma(M_2) = M_1$.*

III. *There are in G only four split circuits, J_1, J_2, J_3, J_4 and only two topologically distinct maps of G on the sphere are possible. The complementary domain boundaries of each map include exactly two split circuits: J_1 and J_2 in one map, and J_3, J_4 in the other map. Any σ is of order 2 and either carries every split circuit into itself or else interchanges both J_1 and J_2, and J_3 and J_4. There are at most four distinct σ's, including the identity.*

Proof. Suppose every homeomorphism of G extendable.

Case 1. $\sigma(J) = J$ for every homeomorphism σ and every split circuit J. If some σ leaves some split circuit J pointwise fixed then σ is the identity.

[11] W. L. Ayres, "Concerning continuous curves and correspondences," *Annals of Mathematics*, vol. 28 (1927), p. 402.

For, map G so that J is a c. d. b. of G, and extend σ to be a periodic homeomorphism σ' of the sphere.[12] Since this σ' has a circuit of fixed points Eilenberg's results[13] concerning periodic transformations of a sphere show that the extension σ' is either the identity or is homeomorphic to a reflection in J. In the latter case σ' would take $G - J$, on one side of J, into something on the other side of J; but there is nothing on the other side of J since J is a c. d. b. of G. Hence σ must be the identity, as asserted.

Consider now any split circuit J and any component N_1 of $G - J$. As in the proof of Lemma 2 choose a split circuit $K = f + F$, where $f \subset N_1$ and $F \subset J$. For any σ, the hypothesis of Case 1 shows that $\sigma(K) = K$, $\sigma(J) = J$ and hence that $\sigma(F) = F$. Let A and B be the ends of the arc F. If A and B are each invariant under σ, then every edge and branch point of J is fixed under σ, and σ is the identity. The only alternative is for σ to interchange A and B.

If there were two homeomorphisms σ and τ, neither the identity, then both σ and τ must interchange A and B. Since $\sigma \cdot \tau^{-1}$ leaves A and B fixed, it therefore will be the identity, so that $\sigma = \tau$. Hence there is, as asserted in I, only one homeomorphism not the identity.

Case 2. For some split circuit J, and some σ; $\sigma(J) \neq J$. By Theorem C there are then only two components N_1 and N_2 of $G - J$. Introduce the subarcs F_1 and F_2 of J and the subarcs f_1 and f_2 of $\sigma(J)$ as in (1) and (2) in the proof of Theorem C, and consider the four split circuits,

$$J,\ \sigma(J),\ F_1 + f_2 + J \cdot \sigma(J),\ f_1 + F_2 + J \cdot \sigma(J), \tag{3}$$

composed of J, f_1 and f_2. We shall show first that if G contains a split circuit K, other than one of (3), then K must lie entirely in one of the sets $H_i = N_i + F_i$ $(i = 1, 2)$.

Suppose G contains a split circuit K that is not one of (3) and has points in both H_1 and H_2. Let g_i be the arc of K that lies in H_i and A and B the ends of F_1. Then the circuits K and L $(= K - g_2 + F_2)$ each pass through A and B. Now one of the arcs, say g_1, must be different from F_1 and from f_1, and L is a split circuit; for any two components of $G - L$ with feet on the arc $L - F_2$ must have feet on distinct arcs (except for end points) of L since

[12] Since G is a graph (consisting of a finite number of vertices and edges) any transformation σ such that $\sigma(G) = G$ is necessarily periodic with a finite period k. Hence, if σ is extendable to the sphere, a transformation σ' of the sphere into itself can be so defined that $\sigma' = \sigma$ for points of G and that σ' is periodic of period k.

[13] S. Eilenberg, "Sur les transformations periodique de la surface de sphere," *Fundamenta Mathematicae* vol. 22 (1934), pp. 28-41.

they are on distinct arcs of the split circuit K, and there is only one component N_2 of $G - L$ with feet on F_2. If G is mapped so that J and $\sigma(J)$ are both c. d. b.'s of the map of G, the region R inside $f_1 + F_1$ (i. e. the region not containing F_2) is cut by the arc g_1. Since $\sigma(J) \neq J$ it follows easily that $\sigma(L) \neq L$, and from Theorem C, $G - L$ has only two components. But since the region R is cut by the arc g_1 of L there must be two components of $G - L$ in $\bar{R}$ and one component N_2 in H_2, making three in all. Therefore, no split circuit except those in (3) has arcs in both H_1 and H_2.

Suppose $\sigma(H_i) = H_i$. Then $\sigma(J) \neq J$, $\sigma(f_i) = F_i$ and $\sigma(F_i) = f_i$. Furthermore, if the end points of F_i remain invariant under σ it is possible to define a homeomorphism τ equal to σ on H_1 and to the identity on $G - H_1$. Then τ is not extendable (Theorem C) contrary to assumption. Hence the end points of F_i are interchanged by σ. Suppose now that H_1 contains a split circuit K such that $\sigma(K) = K$. Then K is in R_1, the "inside" of $f_1 + F_1$. Let R_2 be the region "outside" $f_1 + F_1$. Extend σ to be a periodic transformation σ' of the sphere. Since H_2 lies in R_2, $\sigma'(R_2) = R_2$. The periodic transformation σ' of R_2 must be homeomorphic to a rotation or to reflection [14] of $\bar{R}_2$. The latter is impossible as σ' has no fixed points on the boundary of R_2. The "rotation" σ' thus has a fixed point in R_2, and by Lemma 2 σ' has two fixed points in R_1. But $f_1 + F_1$ is fixed under σ' so that σ' must be homeomorphic to a rotation of the sphere, although a rotation of the sphere cannot have a fixed circuit and three fixed points not on this circuit (Eilenberg). Hence $\sigma(K) \neq K$ so that G contains no split circuit invariant under σ. Therefore, if there is no homeomorphism of G sending H_1 into H_2, σ is the only homeomorphism not the identity. This follows as in the last paragraph of Case 1. Then $\sigma^2 = 1$,[15] and we have condition II.

The only possibility remaining is that some σ has $\sigma(J) \neq J$ with $\sigma(H_1) = H_2$ and $\sigma(H_2) = H_1$. Suppose a split circuit K lies in H_1. Then $\sigma(K) \neq K$ and there are only two components of $G - K$. One of these components must contain H_2, and the other must lie in H_1. Now $\sigma(K)$ must contain an arc in each of these components, and hence must have an arc in H_1. But this is impossible since $\sigma(K)$ must lie in H_2. Therefore, G contains no split circuits other than (3).

Since c. d. b. must go into c. d. b. each of the following six σ's is possible. $\sigma_1(F_i) = f_i$, $\sigma_1(f_i) = F_i$ (as in Case II, interchanging A_i and B_i, where A_i and B_i are end points of F_i); $\sigma_2(F_1) = F_2$, $\sigma_2(F_2) = F_1$, $\sigma_2(f_1) = f_2$, $\sigma_2(f_2) = f_1$ (where A's go into B's and B's into A's) : σ'_2 the same as σ_2, but

[14] Eilenberg, *loc. cit.*

[15] Since σ^2 leaves J pointwise fixed σ^2 is the identity.

A's go into A's and B's into B's; $\sigma_3(F_1) = f_2$, $\sigma_3(F_2) = f_1$, $\sigma_3(f_1) = F_2$, $\sigma_3(f_2) = F_1$ (A's into B's and B's into A's); σ'_3 same as σ_3 but A's go into A's and B's into B's; $\sigma_4(F_i) = F_i$, $\sigma_4(f_i) = f_i$ interchanging A_i and B_i. Now the following pairs cannot occur for the same mapping of G: (σ_1, σ_4), (σ_2, σ_3), (σ'_2, σ'_3). For suppose σ_1 and σ_4 both exist for a given G. Then $\sigma_4\sigma_1(F_i) = f_i$, $\sigma_4\sigma_1(f_i) = F_i$ leaving A_i and B_i fixed. But we have shown above (Case II) that there is no homeomorphism (other than the identity) leaving the end points of F_i fixed and carrying H_i into H_i. A similar contradiction may be obtained for the other two pairs of homeomorphisms. Therefore, there are at most three distinct homeomorphisms (excluding the identity) for any particular mapping of G.

The following relationships hold for the various σ's: $\sigma_2\sigma_1 = \sigma'_3$, $\sigma'_2\sigma_1 = \sigma_3$, $\sigma_2\sigma_4 = \sigma'_2$, $\sigma_3\sigma_4 = \sigma'_3$.

Furthermore, since each of the above σ's is such that σ^2 leaves at least one circuit of fixed points, it follows that $\sigma^2 = 1$ in each case (Eilenberg).

This establishes Case III, and the necessity of the given conditions.

The sufficiency of any one of the above conditions follows from Theorem C.

A further characterization of G when condition I is satisfied may be obtained by using Lemma 2. Let J be a split circuit with the invariant points X and Y. Then J can be expressed as the sum of two arcs, AXB and AYB, so that the feet of any component of $G - J$ lie entirely on AXB or entirely on AYB. Let N_1 be a component of $G - J$ with feet on AXB. Let $AXB = ApXqB$. Then for any point of $\bar{N}_1 \cdot J$ lying on $(ApX - X)$ there is a corresponding point of $\bar{N}_1 \cdot J$ lying on $(XqB - X)$.

In Case II, σ can be shown equivalent to a rotation of the sphere S through 180° about two invariant points, one lying in the region of S bounded by $f_1 + F_1$ that contains H_1, the other in the region of S bounded by $f_2 + F_2$ that contains H_2.

COROLLARY. *If G satisfies the conditions of Theorem D, then G either has no split circuits, or has four split circuits and at most four distinct homeomorphisms, or has at most two homeomorphisms.*

UNIVERSITY OF ARKANSAS,
AND
CORNELL UNIVERSITY.

GEOMETRIC CHARACTERIZATIONS OF INVARIANT PARTIAL DIFFERENTIAL EQUATIONS.*

By Aaron Fialkow.[1]

1. Introduction. This paper deals with a system of curves

$$f(x_1, x_2) = c \tag{1.1}$$

on an oriented two dimensional Riemannian surface element whose metric tensor is[2] g_{ij} so that

$$ds^2 = g_{ij}dx_i dx_j, \qquad |\, g_{ij} \,| > 0. \tag{1.2}$$

It is assumed that $f(x_1, x_2)$ is the solution of certain partial differential equations which involve the components of the metric tensor and their derivatives. The form of the equations which we consider is unchanged by an isometric transformation of the surface or by changes in the coördinate system (x_i). This means that the curves (1.1) which are solutions must have geometric peculiarities which we investigate.

For this purpose, it is useful to construct the orthogonal trajectories of (1.1). Then the geodesic curvatures of the two curves passing through each point and their tangential and normal derivatives are defined and are intrinsic quantities.[3] The intrinsic characterization of the system (1.1) which is subject to various invariant differential conditions is given as relations among these geodesic curvatures and their directional derivatives.[4] In the final sections, we derive equivalent geometric properties of a more synthetic nature if the surface is developable. We also indicate new characterizations of developable and minimal surfaces.

* Received December 14, 1936; revised May 5, 1937.

[1] National Research Council Fellow.

[2] As is customary in tensor analysis, an index which appears twice is to be summed over the range 1, 2.

[3] A detailed study of the intrinsic differential geometry on a surface appears in Graustein, "Méthodes Invariantes dans la Géométrie Infinitésimale des Surfaces," *Mémoires de l'Académie Royale de Belgique* (Classe des Sciences), (2), vol. 11 (1929), and "Invariant methods in classical differential geometry," *Bulletin of the American Mathematical Society*, vol. 36 (1930), pp. 489-521. Cf. in particular the sections on differential invariants.

[4] Similar characterizations are given by the author in *Proceedings of the National Academy of Sciences*, vol. 19 (1933), pp. 543-548.

2. Some invariant relations for an orthogonal net. A single infinitude of curves is given by (1.1). We may introduce new orthogonal curvilinear coördinates

$$u = u(x_1, x_2), \quad v = v(x_1, x_2)$$

so that $v(x_1, x_2) = f(x_1, x_2)$. Then the given orthogonal net is equivalent to the parametric curves

$$v = c \tag{2.1}$$

$$u = c. \tag{2.2}$$

The metric tensor may be written as

$$ds^2 = e^{-2A(u,v)}du^2 + e^{-2B(u,v)}dv^2. \tag{2.3}$$

The geodesic curvature of a curve of (2.1) and its tangential and normal derivatives at a point are represented by γ, γ_s and γ_n. The symbols for the corresponding curve of (2.2) are Γ, Γ_S and Γ_N. An analogous notation is used for higher derivatives with the understanding that $\gamma_{ns} = \dfrac{d}{ds}\dfrac{d\gamma}{dn}$. The arc length of curves of (2.1) and (2.2) is denoted by s and S respectively. The positive tangents and normals are oriented so that

$$ds = -dN, \qquad dn = dS. \tag{2.4}$$

We shall use the following symbols for differential invariants which occur in our work:[5]

$$\begin{aligned}
\Delta_1 f &= g^{ij}f_{,i}f_{,j} \\
\Delta_2 f &= g^{ij}f_{,ij} \\
\Delta_3 f &= g^{ij}g^{lm}[f_{,i}f_{,j}f_{,lm} - f_{,i}f_{,l}f_{,jm}] \\
\Delta_4 f &= \tfrac{1}{2}g^{ij}g^{lm}[f_{,il}f_{,jm} - f_{,ij}f_{,lm}].
\end{aligned}$$

Of course, $\Delta_1 f$ and $\Delta_2 f$ are the well known *first* and *second Beltrami parameters.* We note that since $|g_{ij}| > 0$, $\Delta_1 f > 0$. Hence division by $\Delta_1 f$ is always permissible—a fact we shall use frequently in what follows. When there is no ambiguity we shall write Δ_i for $\Delta_i f$.

In the coördinate system (u, v) having the metric tensor (2.3), the expressions for the geodesic curvatures of (2.1) and (2.2) as well as the Δ's is greatly simplified. This makes it easy to verify[6] that *the following relations hold for any orthogonal net:*

---

[5] The comma in the subscript denotes covariant differentiation as in Eisenhart, *Riemannian Geometry,* p. 26.

[6] Some of the details of the calculations necessary to derive (2.5) to (2.12) are

$$(2.5) \qquad (\Delta_1)_s = -2\Gamma\Delta_1$$

$$(2.6) \qquad \Delta_2 = \tfrac{1}{2}\Delta_1^{-\frac{1}{2}}(\Delta_1)_n - \gamma\Delta_1^{\frac{1}{2}}$$

$$(2.7) \qquad (\Delta_2)_s = -(\gamma_s + \Gamma_s)\Delta_1^{\frac{1}{2}} - 2\Gamma\Delta_2$$

$$(2.8) \qquad \Delta_3 = -\gamma\Delta_1^{3/2}$$

$$(2.9) \qquad (\Delta_3)_s = (3\gamma\Gamma - \gamma_s)\Delta_1^{3/2}$$

$$(2.10) \qquad \Delta_4 = \tfrac{1}{2}(\gamma(\Delta_1)_n - \Gamma(\Delta_1)_s).$$

The Gauss equation is equivalent to[7]

$$(2.11) \qquad K = \gamma_n + \Gamma_N - \gamma^2 - \Gamma^2$$

where K is the Gaussian curvature of the surface. Every point function $h(x_1, x_2)$ satisfies the integrability condition[8]

$$(2.12) \qquad h_{sn} - h_{ns} = \gamma h_s + \Gamma h_n.$$

These equations are useful in the characterization of the solutions of various invariant types of differential equations.

We first prove that an analytic orthogonal net is completely determined by the values of γ, Γ and their derivatives at one point. By the classical theory, the values of

$$(2.13) \qquad \gamma, \gamma_s, \gamma_{ss}, \cdots$$

and

$$(2.14) \qquad \Gamma, \Gamma_S, \Gamma_{SS}, \cdots$$

determine the curves of (2.1) and (2.2) respectively passing through the point if the initial direction of one of them is given. We refer to these curves as the base curves. It will suffice to prove that through each point of the base curve of one family which is sufficiently near the given point there passes a unique curve of the orthogonal family. At any point of the base curve of (2.2), the corresponding curve of (2.1) depends on the values of the quantities (2.13) at this point. But these quantities are determined at any point of the base curve of (2.2) by equations of which

given in a forthcoming paper which characterizes the solutions of $\Delta_2 f = F(f)$ and $\Delta_2 f = af + b$ and gives some physical applications.

[7] Cf. Darboux, "Lecons Sur La Théorie Générale des Surfaces," vol. 3 (1894), p. 131.

[8] This condition is a special case of general integrability conditions which for an n-dimensional Riemannian space are due to Ricci. Cf. Eisenhart, *loc. cit.*, p. 99 and Graustein. *Bulletin*, *loc. cit.*, pp. 497-499.

$$\gamma = (\gamma)_0 + (\gamma_n)_0 \cdot S + \frac{1}{2!} (\gamma_{nn})_0 \cdot S^2 + \cdots$$
$$\gamma_s = (\gamma_s)_0 \cdot (\gamma_{sn})_0 \cdot S + \frac{1}{2!} (\gamma_{snn})_0 \cdot S^2 + \cdots$$

are the first two. Thus the corresponding curve of (2.1) is fixed. In the same way, we show that through each point of the base curve of (2.1), there passes a determinate curve of (2.2). This proves

THEOREM I. *The curves of an analytic orthogonal net are completely determined, except for initial direction, by the values of the geometric quantities* (2.13) *and* (2.14) *and their intrinsic normal derivatives of all orders at a single point.*

3. Geometric characterization of invariant orthogonal nets. In this section, we shall assume that $f(x_1, x_2)$ is the solution of an invariant partial differential equation and obtain characteristic properties of the corresponding orthogonal net. Such properties are known for some equations. Thus if $\Delta_1 f = F(f)$, then the curves (1.1) are parallel, i.e., $\Gamma = 0$. This well known result follows at once from (2.5) since $(\Delta_1)_s = F'(f) \cdot f_s$ and f_s is always identically zero. A characteristic property for the solutions of Laplace's equation, $\Delta_2 f = 0$, is also known.[9] It is $\gamma_s + \Gamma_s = 0$ and may easily be deduced from (2.7).

The two geometric equations given above involve the geodesic curvatures and their directional derivatives and are intrinsic quantities of the orthogonal net. On the other hand, the differential invariants although independent of the choice of the coördinate system (x_i) are not intrinsic since they depend upon the particular parameter f used to describe the curves (1.1). In what follows, we shall convert the non-intrinsic description of a system of curves by means of differential invariants into an intrinsic one by the use of geodesic curvatures. Valued suggestions by Dr. Harry Levy resulted in a simplification of our original proofs.

We now assume that

$$\Delta_4 f = F(f) \tag{3.1}$$

and discuss the geometry of the corresponding orthogonal net. From (3.1) it follows that $(\Delta_4)_s = 0$ since $f_s = 0$. We differentiate (2.10) with respect to s and set the right-hand member equal to zero. This gives

$$\gamma_s(\Delta_1)_n + \gamma(\Delta_1)_{ns} - \Gamma_s(\Delta_1)_s - \Gamma(\Delta_1)_{ss} = 0. \tag{3.2}$$

[9] Cf. Darboux, *loc. cit.*, vol. 3, p. 154.

From (2.12),

$$(\Delta_1)_{ns} = (\Delta_1)_{sn} - \gamma(\Delta_1)_s - \Gamma(\Delta_1)_n. \tag{3.3}$$

Substituting this expression in (3.2) and using (2.4) and (2.5) we obtain

$$(\Delta_1)_n[\gamma_s - 3\gamma\Gamma] - 2\Delta_1[\gamma\Gamma_S + 2\Gamma\Gamma_N - \gamma^2\Gamma + 2\Gamma^3] = 0. \tag{3.4}$$

If

$$\gamma_s - 3\gamma\Gamma = 0 \tag{3.5}$$

then it follows that

$$\gamma\Gamma_S + 2\Gamma\Gamma_N - \gamma^2\Gamma + 2\Gamma^3 = 0. \tag{3.6}$$

Conversely, if (3.5) and (3.6) are satisfied, (3.2) is true and (3.1) follows for some $F(f)$. Indeed it is easy to show that in this case if γ and Γ do not both vanish, if F is given, a suitable parameter $g(f)$ may be found so that

$$\Delta_4 g = F(g). \tag{3.7}$$

For $g(f) = c$ is the same system of curves as (1.1) and

$$\Delta_4 g(f) = g'(f)[g''(f)\cdot\Delta_3 f + g'(f)\cdot\Delta_4 f]. \tag{3.8}$$

From (3.5) and (2.9), $(\Delta_3 f)_s = 0$ or $\Delta_3 f = h(f)$. Since $\Delta_4 f = k(f)$ it follows from (3.8) that (3.7) has solutions $g(f)$ for all $F(g)$. The only exception occurs when $\Delta_3 f = \Delta_4 f = 0$. From (2.8) and (2.10) this means that $\gamma = \Gamma = 0$.

If $\gamma_s - 3\gamma\Gamma \neq 0$, we consider the system of equations consisting of (2.5) and (3.4). The condition of integrability of this system is (3.3). This becomes after simplifying by means of (2.5) and (3.4),

$$(\gamma_s - 3\gamma\Gamma)[(\gamma\Gamma_S + 2\Gamma\Gamma_N - \gamma^2\Gamma + 2\Gamma^3)_s + \Gamma(\gamma\Gamma_S + 2\Gamma\Gamma_N - \gamma^2\Gamma + 2\Gamma^3) + (\gamma_s - 3\gamma\Gamma)(\Gamma_S - \gamma\Gamma)] - (\gamma\Gamma_S + 2\Gamma\Gamma_N - \gamma^2\Gamma + 2\Gamma^3)(\gamma_s - 3\gamma\Gamma)_s = 0.$$

This proves

THEOREM II. *The necessary and sufficient condition that a one parameter family of curves on a surface, $f(x_1, x_2) = c$, be a solution of $\Delta_4 f = F(f)$ is that either $\gamma_s - 3\gamma\Gamma \neq 0$ and*

$$\begin{aligned}(\gamma_s - 3\gamma\Gamma)[&(\gamma\Gamma_S + 2\Gamma\Gamma_N - \gamma^2\Gamma + 2\Gamma^3)_s \\ &+ \Gamma(\gamma\Gamma_S + 2\Gamma\Gamma_N - \gamma^2\Gamma + 2\Gamma^3) + (\gamma_s - 3\gamma\Gamma)(\Gamma_S - \gamma\Gamma)] \\ &- (\gamma\Gamma_S + 2\Gamma\Gamma_N - \gamma^2\Gamma + 2\Gamma^3)(\gamma_s - 3\gamma\Gamma)_s = 0 \qquad (Type\ I)\end{aligned}$$

or that

$$\gamma_s - 3\gamma\Gamma = 0$$

and $$\gamma\Gamma_S + 2\Gamma\Gamma_N - \gamma^2\Gamma + 2\Gamma^3 = 0. \qquad (\textit{Type II})$$

Any particular family of Type II which does not consist of parallel geodesics can be parametered so that it is a solution of $\Delta_4 f = F(f)$ *for every* $F(f)$.

By similar methods, the condition that $F(f)$ be constant may be obtained. By the last part of Theorem II, it is only necessary to consider families of Type I. In addition to the above property of Type I, we find another second order equation which guarantees that $F(f)$ is a constant. As this equation is complicated, we do not give it explicitly here. However, there is a simple characterization when

$$\Delta_4 f = 0 \tag{3.9}$$

which is derived below. These curves are related to developable surfaces in a manner which is considered in § 5.

If $\gamma \not\equiv 0$, from (3.9), (2.5) and (2.10) we have

$$(\Delta_1)_n = -\frac{2\Gamma^2}{\gamma}\Delta_1. \tag{3.10}$$

This equation and (2.5) form a system whose integrability condition is (3.3). When this is simplified by means of (2.4), (2.5) and (3.10) we obtain

$$2\gamma\Gamma\Gamma_N + \Gamma^2\gamma_s + \gamma^2\Gamma_S - \gamma^3\Gamma - \gamma\Gamma^3 = 0.$$

It follows from (2.11) that this is equivalent to

$$\Gamma^2\gamma_s + \gamma\Gamma(\Gamma_N - \gamma_n) + \gamma^2\Gamma_S + \gamma\Gamma K = 0.$$

This result is stated in

THEOREM III. *The necessary and sufficient condition that a one-parameter family of curves on a surface,* $f(x_1, x_2) = c$, *which does not consist entirely of geodesics be a solution of* $\Delta_4 f = 0$ *is that*

$$\Gamma^2\gamma_s + \gamma\Gamma(\Gamma_N - \gamma_n) + \gamma^2\Gamma_S + \gamma\Gamma K = 0.$$

If $\gamma \equiv 0$, all the curves are geodesics. If they are solutions of (3.9), it follows from (2.5) and (2.10) that $\Gamma \equiv 0$, i.e., the geodesics are parallel. From (2.11), this can occur only on a developable surface. This shows that *the only systems of geodesics,* $f(x_1, x_2) = c$, *satisfying* $\Delta_4 f = 0$ *are families of parallel geodesics on a developable surface.*

More generally, if a one-parameter family of geodesics is a solution of (3.1), it is of Type II and according to Theorem II, $\Gamma(\Gamma^2 + \Gamma_N)' = 0$. From

(2.11) and this equation we find that *the only systems of geodesics,* $f(x_1, x_2) = c$, *which are solutions of* $\Delta_4 f = F(f)$ *are such that* $\Gamma = \sqrt{-(K/2)}$. In particular, there cannot be a family of geodesics satisfying (3.1) on a region of a surface where the Gaussian curvature is positive.

We now determine the characteristic properties of the orthogonal net if

$$(\Delta_2 + \Delta_3) f = 0. \tag{3.11}$$

These curves are related to minimal surfaces as is shown in § 5. From (2.6) and (2.8) we find that (3.11) may be written as

$$(\sqrt{\Delta_1})_n = \gamma\sqrt{\Delta_1} + \gamma(\sqrt{\Delta_1})^3. \tag{3.12}$$

The identity (2.5) is equivalent to

$$(\sqrt{\Delta_1})_s = -\Gamma\sqrt{\Delta_1}. \tag{3.13}$$

The integrability condition of (3.12) and (3.13) is (2.12) where $h = \sqrt{\Delta_1}$. This condition, after simplification may be written as

$$(\gamma_s + \Gamma_S) + (\gamma_s - \gamma\Gamma)\Delta_1 = 0. \tag{3.14}$$

This means that the geometric characterization of (3.11) is obtained either by equating the coefficients of 1 and Δ_1 in (3.14) to zero or by substituting the value of Δ_1 given by (3.14) in both (3.12) and (3.13). In the first case the family is isothermal. This completes the proof of

THEOREM IV. *The necessary and sufficient conditions that a one-parameter family of curves on a surface,* $f(x_1, x_2) = c$, *be a solution of* $(\Delta_2 + \Delta_3) f = 0$ *is that either the family be non-isothermal and*

$$\left[\frac{\gamma_s + \Gamma_S}{\gamma_s - \gamma\Gamma}\right]_s + 2\Gamma\left[\frac{\gamma_s + \Gamma_S}{\gamma_s - \gamma\Gamma}\right] = 0$$

and

$$\left[\frac{\gamma_s + \Gamma_S}{\gamma_s - \gamma\Gamma}\right]_n - 2\gamma\left[\frac{\gamma_s + \Gamma_S}{\gamma_s - \gamma\Gamma}\right] + 2\gamma\left[\frac{\gamma_s + \Gamma_S}{\gamma_s - \gamma\Gamma}\right]^2 = 0$$

or that

$$\gamma_s + \Gamma_S = 0 \quad \textit{and} \quad \gamma_s - \gamma\Gamma = 0.$$

The geometric description of orthogonal nets which are solutions of $\Delta_3 f = F(f)$ follows immediately from (2.9) by setting the right-hand member equal to zero. Hence $\gamma_s - 3\gamma\Gamma = 0$ *is the characteristic property of the solutions of* $\Delta_3 f = F(f)$. Similarly, it follows from (2.8) that *any one-parameter system of geodesics is a solution of* $\Delta_3 f = 0$ *and conversely.*

4. Equivalent geometric properties on a developable surface. We consider some equivalent characterizations of the invariant systems of curves discussed in the previous section under the assumption that they are imbedded in a developable surface. In what follows, it is understood that the customary terminology of plane figures is to be applied to their isometric generalizations on any developable surface. The curves are referred to a system of coördinates (x, y) which is isometrically equivalent to plane Cartesian coördinates.

The characteristic property of curves which are solutions of $\Delta_3 f = F(f)$ as noted at the end of § 3 is

$$\gamma_s - 3\gamma\Gamma = 0. \tag{4.1}$$

The axis of deviation of a curve at a point is defined as follows: Chords are drawn parallel to the tangents at the point and terminating in the curve. The limiting position of the geodesic joining the midpoint of each chord to the point is the axis of deviation. Thus the axis is completely determined by the limiting position of four points which approach coincidence. Two of these determine the tangent; the other two fix a parallel chord. Hence the expression for the angle between the axis of deviation and the tangent involves only the first three derivatives. From analytic geometry, we find that if ϕ is the angle between the axis and the tangent geodesic

$$\tan\phi = -\frac{3\gamma^2}{\gamma_s}. \tag{4.2}$$

It is well known property of the parabola that the locus of the midpoints of parallel chords is a line parallel to the principal axis. Since the osculating parabola of a curve has third order contact with the curve, it follows that the axis of deviation at a point of a curve may also be defined as the geodesic through the point which is parallel to the principal axis of the corresponding osculating parabola.

The locus of points where the slope of the curves (1.1) is constant is called the system of *isoclines* or *isoclinal curves* of (1.1).[10] The equation of the isoclines is $f_x(x, y) + cf_y(x, y) = 0$ where c is an arbitrary constant. It is easy to show that the angle θ between a curve of (1.1) and the corresponding isocline is determined by

$$\tan\theta = -\frac{\gamma}{\Gamma}. \tag{4.3}$$

[10] The isogonal trajectories of a one-parameter family of curves consists of those curves which cut the family at a constant angle α. Each family obtained for a fixed value of α is called a base family of the system of isogonal trajectories. It is clear that the same isoclines are obtained for any base family. Thus the isoclines are a property of the whole isogonal system.

From (4.2) and (4.3) it follows that (4.1) is equivalent to $\phi = \theta$. This proves

THEOREM V. *The characteristic property of a one-parameter family of curves on a developable surfrace, $f(x_1, x_2) = c$, which is a solution of $\Delta_3 f = F(f)$, $(F \not\equiv 0)$ is that at each point the axis of deviation is tangent to the isoclinal curve.*

Straightforward calculation shows that the curvature of an isoclinal curve is

$$-\frac{\gamma^2\Gamma_S + \gamma\Gamma(\Gamma_N - \gamma_n) + \Gamma^2\gamma_s}{(\gamma^2 + \Gamma^2)^{3/2}}.$$

According to Theorem III, since $K = 0$, the characteristic property of curves which are not geodesics and which are solutions of $\Delta_4 f = 0$ is the vanishing of the numerator of this expression. Hence the property is equivalent to the condition that the isoclines be geodesics.[11] A family of curves of this kind is completely determined by any one-parameter family of geodesics and a single curve which cuts them. As a result, we have

THEOREM VI. *The only one-parameter families of curves on a developable surface, $f(x_1, x_2) = c$, which are solutions of $\Delta_4 f = 0$ are families of*

(1) *parallel geodesics*

(2) *curves whose isoclines are geodesics.*

If a system of curves is a solution of $\Delta_4 f = F(f)$ for every $F(f)$, it is of Type II. When $K = 0$, the second of the characteristic equations for this type given by Theorem II may be transformed by means of the first and (2.11) into

$$\Gamma^2\gamma_s + \gamma\Gamma(\Gamma_N - \gamma_n) + \gamma^2\Gamma_N = 0.$$

Hence by the results obtained in this section, the isoclines and the axes of deviation of the family must form the same system of geodesics. Since the isoclinal curves are geodesics, the differential equation of the family must be

$$y = x \cdot f(p) + g(p) \tag{4.4}$$

where $p = dy/dx$. Here $f(p)$ is the slope of an isocline on which the curves have slope p. Since the isocline is also the axis of deviation, the angle ϕ between the axis and the tangent geodesic is given by

---

[11] Since this condition also refers to the whole isogonal system, it follows that if one base family is a solution of $\Delta_4 f = 0$ then all are which do not consist entirely of geodesics.

$$\tan\phi = \frac{f(p)-p}{1+p\cdot f(p)}. \tag{4.5}$$

A comparison of (4.2) and (4.5) leads to

$$\frac{y'''(1+p^2)-3py''^2}{y''^2} = \frac{3(1+p\cdot f(p))}{p-f(p)} \tag{4.6}$$

where γ and γ_s are represented by their equivalent expressions in the derivatives of y. Hence any family of Type II must be a common solution of (4.4) and (4.6).

After some discussion, we find that for the solutions of these two equations, either

$$f(p) = c; \qquad g(p) = \frac{a}{p} + b \tag{4.7}$$

or

$$f(p) = \frac{b-ap}{a+p}; \qquad g(p) = cf + d. \tag{4.8}$$

The solution of (4.4) and (4.7) is found to be any family of congruent parabolas with the same principal axis. The curves satisfying (4.4) and (4.8) form any system of similar ellipses or hyperbolas. That any family of similar ellipses or hyperbolas is of Type II follows at once, since it is a well known property of these curves that a diameter drawn to a point bisects the chords parallel to the tangent at that point. Of course, an analogous statement holds for the parabolas. This proves

THEOREM VII. *The only families of curves on a developable surface, $f(x_1, x_2) = c$, which are solutions of $\Delta_4 f = F(f)$ for every $F(f)$ are families of similar ellipses or hyperbolas and families generated by the translation of an arbitrary parabola parallel to its principal axis.*

5. Applications to the geometry of surfaces. Let an arbitrary surface be cut by all planes parallel to a fixed plane. If the single infinitude of plane sections is projected orthogonally upon the fixed plane, a one-parameter family of plane curves is obtained.[12] The following questions suggest themselves: What are the necessary geometric properties of this family if the surface is known to be of a given type defined by an invariant differential condition? Conversely, what properties of a one-parameter family of plane curves are

[12] We exclude the trivial case in which the family degenerates into a single curve. This occurs if a cylindrical surface is cut by planes perpendicular to its generators.

sufficient to guarantee a surface of this type such that the given curves are the projections of its parallel plane sections? The answers to these questions follow readily from our previous work in the case of developable and minimal surfaces. The proofs indicate how analogous results may be obtained for other types of surfaces.

Let

$$z = f(x, y) \tag{5.1}$$

be the equation of a surface which is cut by ∞^1 parallel planes

$$z = c. \tag{5.2}$$

Then the projections of the intersections of (5.2) with (5.1) are the family of plane curves (1.1). Since in the plane, $\Delta_4 f$ is the Hessian $f_{xy}{}^2 - f_{xx}f_{yy}$, the surface (5.1) is developable if and only if $\Delta_4 f = 0$. The conclusion of Theorem VI relative to this type of family enables us to characterize it geometrically. The necessary property of these curves may also be deduced directly from the character of the surface in a simple manner. A developable surface may be considered as generated by the tangent lines of a curve in space. The curve is the edge of regression of the surface. The tangent to a plane section at any point is determined by the intersection of the tangent plane of the surface with the plane of the section. But the tangent plane is the same at any point of a generator. Hence the angle between a fixed generator and the tangent to any plane section at a point where it meets this generator is constant. Therefore the projections of the generators are the isoclinal straight lines of the family of projected plane sections. The projection of the edge of regression is clearly the envelope of these straight lines.

In the degenerate case in which the family consists of straight lines, the plane sections on the surface must also be straight lines. Since the lines must be parallel the surface is a cylinder.

We note that if one function $f(x, y)$ is known which is a solution of $\Delta_4 f = 0$, the most general function is a solution of the equation obtained by setting the right-hand member of (3.8) equal to zero. If $\Delta_3 f \neq 0$, that is, if $\gamma \neq 0$, it follows that the most general function is $af + b$. This proves

THEOREM VIII. *The necessary and sufficient condition that a one-parameter family of plane curves that are not all straight lines be the projections of the ∞^1 parallel plane sections of a developable surface is that their isoclines be straight lines. The isoclines and their envelope are the projections of the*

generators and the edge of regression of the surface respectively. Any surface associated with a fixed family of curves in a fixed plane may be obtained from one such surface by a linear transformation orthogonal to the fixed plane. If the family of curves is composed entirely of straight lines, they must be parallel. In this case, the straight lines are the projections of the parallel generators of a cylindrical surface.

The surface (5.1) is minimal if and only if $(\Delta_2 + \Delta_3)f = 0$ since in the plane, $(\Delta_2 + \Delta_3)f = (f_y^2 + 1)f_{xx} - 2f_xf_yf_{xy} + (f_x^2 + 1)f_{yy}$. Applying the results of Theorem IV, we obtain a characterization of minimal surfaces by means of the geometric properties of their parallel plane sections similar to that given above for developable surfaces.

PRINCETON UNIVERSITY
AND THE
INSTITUTE FOR ADVANCED STUDY.

SETS OF CONJUGATE MATRICES.*

By E. T. Browne.

1. **Introduction.** Let λ be an indeterminate scalar and let M be any given n-square matrix with elements independent of λ. If then $\phi(\lambda)$ is a scalar polynomial, and $M_1, M_2, \cdots, M_{\nu-1}$ are $\nu - 1$ matrices satisfying the conditions:

I. The matrices M_j $(j = 0, 1, \cdots, \nu - 1)$ are commutative in pairs;

II. The condition

$$(1) \qquad (\lambda - M)(\lambda - M_1) \cdots (\lambda - M_{\nu-1}) \equiv \phi(\lambda)^1$$

is satisfied identically in λ; we shall say that the $\nu - 1$ matrices $M_1, \cdots, M_{\nu-1}$ constitute a set of conjugates to M.

From (1) it is clear that the degree of the polynomial $\phi(\lambda)$ is ν. If ν is the *smallest* positive integer such that an identity of the type (1) subsists, we shall say that the $\nu - 1$ matrices M_j constitute a *reduced* set of conjugates to M. In the contrary case, we shall say that these matrices constitute an *extended* set of conjugates to M. From a reduced set of conjugates, one obvious way in which we can get an extended set is by adjoining to the former an arbitrary number of scalar matrices.

If in (1) we replace the scalar indeterminate λ by M, we see at once that $\phi(M) = 0$. Hence, $\phi(\lambda)$ must be divisible by the reduced characteristic function of M. However, it is known, and it will be clear from the results of this paper, that it suffices to take $\phi(\lambda)$ as the reduced characteristic function of M, so that henceforth in this paper $\phi(\lambda)$ will be used in precisely that sense.

The notion of a set of matrices conjugate to a given matrix was introduced by Taber,[2] who, however, confined his attention to matrices M of the third order with *distinct* characteristic roots. In this particular case, the reduced characteristic function coincides with the characteristic function $f(\lambda)$ of M. Taber imposed a third condition, *viz.*,

III. Each M_j has $f(\lambda)$ as its characteristic function.

In 1921 P. Franklin[3] employed the function $f(\lambda)$ on the right in (1)

* Received February 3, 1937.

[1] Here and elsewhere throughout this paper, the symbols λ and $\phi(\lambda)$ are employed to denote the scalar matrices λI, $\phi(\lambda)I$, where I denotes the unit matrix.

[2] Taber, *American Journal of Mathematics*, vol. 13 (1891), pp. 159-172.

[3] Franklin, *Annals of Mathematics*, 2nd Ser., vol. 23 (1923), pp. 97-100.

and showed that by omitting the condition III, a set of generalized conjugates could be set up for any square matrix M. E. S. Sokolnikoff[4] in 1933 was the first to employ the function $\phi(\lambda)$ in (1) and she was led thereby to a consideration of "matrices conjugate to a given matrix with respect to its minimum equation."

We also employ the reduced characteristic function $\phi(\lambda)$ on the right in (1) and it is the purpose of this paper to give an *à priori* method for deriving, along with other sets, the most general sets of conjugates which are expressible as polynomials in M. Before proceeding to the particular problem, however, we can make the following observation. If a set of $\nu-1$ matrices exists satisfying I and II, and Sokolnikoff has shown that such sets do exist, we have on putting $\lambda = M_j$ in (1)

$$\phi(M_j) = 0, \qquad (j = 1, \cdots, \nu-1).$$

We therefore have

THEOREM I.[5] *If $M_1, \cdots, M_{\nu-1}$ constitute a reduced set of conjugates to a matrix M, the reduced characteristic function of each M_j is a factor of the reduced characteristic function of M.*

Moreover, from the identity (1) we see at once that the elementary symmetric functions of the matrices $M, M_1, \cdots, M_{\nu-1}$ are equal to the elementary symmetric functions of the roots of $\phi(\lambda) = 0$, these latter functions being considered as scalar matrices.

2. The case in which M has a single characteristic root. We consider first the case in which M has a single characteristic root α, so that the reduced characteristic function of M reduces to

$$(2) \qquad \phi(\lambda) = (\lambda - \alpha)^{\nu} \qquad (\nu \leqq n).$$

If we write

$$(3) \qquad M = \alpha + \eta$$

i. e., $M - \alpha = \eta$, it is obvious from (2) that η is nilpotent of index ν, and that the matrices

$$(4) \qquad 1, \eta, \eta^2, \cdots, \eta^{\nu-1}$$

---

[4] Sokolnikoff, *American Journal of Mathematics*, vol. 55 (1933), pp. 167-180.

For a complete bibliography up to 1933 see MacDuffee, *The Theory of Matrices*, Berlin (Springer), 1933. Cf. also Hermann, "Über Matrixgleichungen und die Zerlegung von Polynomen in Linearfaktoren," *Compositio Mathematica*, vol. 1 (1934), pp. 284-302; Richardson, "Conjugate matrices," *Quarterly Journal of Mathematics*, vol. 7 (1936), pp. 256-270.

[5] Sokolnikoff, *loc. cit.*, p. 173.

are linearly independent. Moreover any polynomial in M can be written in the form

(5) $$M_j = \alpha_j + \beta_j\eta + \gamma_j\eta^2 + \cdots + \kappa_j\eta^{\nu-1},$$

and conversely, any matrix of the form M_j in (5) is a polynomial in M.

In this case (1) becomes

(6) $$(\lambda - M)(\lambda - M_1)\cdots(\lambda - M_{\nu-1}) \equiv (\lambda - \alpha)^\nu,$$

and our first problem is to determine matrices M_j which are polynomials in M and which satisfy (6).

Since each M_j is expressible in the form (5), the left member of (6) is apparently a polynomial $F(\lambda, \eta)$ in λ and η. But the right member is free of η so that the left member must be also. Since λ^ν is the reduced characteristic function of η, and since in this connection η behaves in all ways precisely as a scalar, save that $\eta^\nu = 0$, this last condition is possible if, and only if, $F(\lambda, \eta)$ is divisible by η^ν; i. e., if and only if F and its first $\nu - 1$ derivatives as to η vanish for $\eta = 0$. Putting $\eta = 0$ in (6) we obtain as a first necessary condition

$$F(\lambda, 0) = (\lambda - \alpha)(\lambda - \alpha_1)\cdots(\lambda - \alpha_{\nu-1}) \equiv (\lambda - \alpha)^\nu,$$

identically in λ. From this we conclude that

(7) $$\alpha = \alpha_1 = \alpha_2 = \cdots = \alpha_{\nu-1}.$$

Now designate by $M_j(\eta)$ the function on the right in (5) and differentiate $F(\lambda, \eta)$ partially as to η. We get

$$\frac{\partial F}{\partial \eta} = \sum_{j=0}^{\nu-1} \frac{\partial F}{\partial M_j} \cdot \frac{\partial M_j}{\partial \eta} = -\sum \frac{F(\lambda, \eta)}{\lambda - M_j} \cdot \frac{\partial M_j}{\partial \eta} = -(\lambda - \alpha)^\nu \sum \frac{\dfrac{\partial M_j}{\partial \eta}}{\lambda - M_j}.$$

Similarly,

$$\frac{\partial^2 F}{\partial \eta^2} = -(\lambda - \alpha)^\nu \sum \left\{ \frac{\dfrac{\partial^2 M_j}{\partial \eta^2}}{\lambda - M_j} + \frac{\left(\dfrac{\partial M_j}{\partial \eta}\right)^2}{(\lambda - M_j)^2} \right\};$$

$$\frac{\partial^3 F}{\partial \eta^3} = -(\lambda - \alpha)^\nu \sum \left\{ \frac{\dfrac{\partial^3 M_j}{\partial \eta^3}}{\lambda - M_j} + \frac{3\dfrac{\partial M_j}{\partial \eta} \cdot \dfrac{\partial^2 M_j}{\partial \eta^2}}{(\lambda - M_j)^2} + \frac{2\left(\dfrac{\partial M_j}{\partial \eta}\right)^3}{(\lambda - M_j)^3} \right\};$$

and in general for $k = 4, 5, \cdots, \nu - 1$,

$$\frac{\partial^k F}{\partial \eta^k} = -(\lambda - \alpha)^\nu \sum \left\{ \frac{\dfrac{\partial^k M_j}{\partial \eta^k}}{\lambda - M_j} + \cdots + \frac{(k-1)!\left(\dfrac{\partial M_j}{\partial \eta}\right)^k}{(\lambda - M_j)^k} \right\},$$

where the dots indicate terms whose denominators are $(\lambda - M_j)^2, \cdots,$

$(\lambda - M_j)^{k-1}$. Since $(\lambda - \alpha)^\nu \not\equiv 0$, we have on putting $\eta = 0$, the following identities in λ:

$$(8) \qquad \sum \frac{\beta_j}{\lambda - \alpha_j} \equiv 0,$$

$$(9) \qquad \sum \left\{ \frac{2\gamma_j}{\lambda - \alpha_j} + \frac{\beta_j^2}{(\lambda - \alpha_j)^2} \right\} \equiv 0,$$

$$(10) \qquad \sum \left\{ \frac{6\delta_j}{\lambda - \alpha_j} + \frac{6\beta_j\gamma_j}{(\lambda - \alpha_j)^2} + \frac{2\beta_j^3}{(\lambda - \alpha_j)^3} \right\} \equiv 0,$$

$$(11) \qquad \sum \left\{ \frac{24\epsilon_j}{\lambda - \alpha_j} + \frac{24\beta_j\delta_j + 12\gamma_j^2}{(\lambda - \alpha_j)^2} + \frac{24\beta_j^2\gamma_j}{(\lambda - \alpha_j)^3} + \frac{6\beta_j^4}{(\lambda - \alpha_j)^4} \right\} \equiv 0,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$(12) \qquad \sum \left\{ \frac{\nu!\,\kappa_j}{\lambda - \alpha_j} + \cdots + \frac{(\nu - 1)!\,\beta_j^{\nu-1}}{(\lambda - \alpha_j)^{\nu-1}} \right\} \equiv 0.$$

Since the α's are all equal, we conclude at once from (8) that

$$(13) \qquad \sum \beta_j = 0.$$

If we rewrite (9) in the form

$$2(\lambda - \alpha) \sum \gamma_j + \sum \beta_j^2 \equiv 0,$$

we see at once that also

$$(14) \qquad \sum \beta_j^2 = 0, \qquad \sum \gamma_j = 0;$$

and similarly from (10), (11) and (12)

$$(15) \qquad \sum \delta_j = \sum \epsilon_j = \cdots = \sum \kappa_j = 0,$$

$$(16) \qquad \sum \beta_j\gamma_j = \sum (2\beta_j\delta_j + \gamma_j^2) = \sum \beta_j^2\gamma_j = 0,$$

$$(17) \qquad \sum \beta_j^3 = \sum \beta_j^4 = \cdots = \sum \beta_j^{\nu-1} = 0.$$

Not only are these conditions necessary but they are sufficient that matrices M_j given by (5) shall satisfy the condition (6).

In view of (13), (14) and (17) it is obvious from Newton's identities that the β's are roots of an equation of the type

$$x^\nu - \beta^\nu = 0,$$

whence, since $\beta = 1$, the remaining β's are uniquely determined save for order. Indeed their values are

$$\omega, \omega^2, \cdots, \omega^{\nu-1}$$

where ω is a primitive ν-th root of unity.

The remaining conditions (14), $\cdots$, (16) are obviously satisfied if we take

$$\gamma_j = \delta_j = \epsilon_j = \cdots = \kappa_j = 0.$$

In this case we have as a set of conjugates to the matrix $M = \alpha + \eta$ the matrices

$$M_j = \alpha + \omega^j \eta \qquad (j = 1, \cdots, \nu - 1). \tag{18}$$

These conjugates were given first by P. Franklin, and were called by him *generalized* conjugates. They are obviously polynomials in M, and have not only the same characteristic function but the same elementary divisors as M.

However, there are other solutions than these for the γ's, δ's, etc. Indeed, since $\gamma = 0$, the ratios of the remaining γ's are determined uniquely from the $\nu - 2$ homogeneous linear equations

$$\begin{aligned} &\sum_1^{\nu-1} \gamma_j = 0, \\ &\sum \beta_j \gamma_j = 0, \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ &\sum \beta_j^{\nu-3} \gamma_j = 0. \end{aligned} \tag{19}$$

Then, since the β's are distinct, the δ's are determined (but not uniquely) by the non-homogeneous equations

$$\begin{aligned} \sum_1^{\nu-1} \delta_j &= 0, \\ 2 \sum \beta_j \delta_j &= - \sum \gamma_j^2, \\ \cdot \quad \cdot \quad \cdot \quad &\cdot \quad \cdot \quad \cdot \end{aligned} \tag{20}$$

etc.

For $\nu = 3$, we have as the most general set of conjugates which are expressible as polynomials in M

$$M = \alpha + \eta, \qquad M_1 = \alpha + \omega\eta + \gamma\eta^2, \qquad M_2 = \alpha + \omega^2\eta - \gamma\eta^2, \tag{21}$$

where ω is a complex cube root of unity and γ is arbitrary.

For $\nu = 4$, the β's are $1, i, -1, -i$, where $i = \sqrt{-1}$. Then from the relations

$$\begin{aligned} \gamma_1 + \gamma_2 + \gamma_3 &= 0, \\ i\gamma_1 - \gamma_2 - i\gamma_3 &= 0, \end{aligned}$$

the γ's are determined to be

$$\gamma_1 : \gamma_2 : \gamma_3 = 1 : i - 1 : -i,$$

while the δ's are subject to the single condition

$$\delta_1 + \delta_2 + \delta_3 = 0.$$

For $\nu = 4$, we have therefore the set of conjugates

$$\begin{aligned} M &= \alpha + \eta, \\ M_1 &= \alpha + i\eta + \rho\eta^2 + \delta_1\eta^3, \\ M_2 &= \alpha - \eta + \rho(i - 1)\eta^2 + \delta_2\eta^3, \\ M_3 &= \alpha - i\eta - \rho i\eta^2 - (\delta_1 + \delta_2)\eta^3. \end{aligned} \tag{22}$$

We obtain in this way all reduced sets of conjugates to M which are expressible as polynomials in M. If $\nu = n$, i. e., if M has a single elementary divisor $(\lambda - \alpha)^n$, the only matrices commutative with M are polynomials in M. In this case the foregoing method yields *all* reduced sets of conjugates to M.

It should be noted particularly that the values of the β's, γ's, δ's, etc., depend in no wise on the value of α, but are dependent on ν only.

We now state a theorem which may be looked upon as essentially a corollary to Theorem I:

THEOREM II. *If M is a matrix of the form $\alpha + \eta$, where η is nilpotent of index ν, there does not exist a set of less than ν matrices, one of which is M itself, such that an identity of the type* (1) *subsists.*

3. Cyclic sets of conjugates. Referring to the matrix

$$M = \alpha + \eta$$

of the preceding section, we know that the matrix

$$M_1 = \alpha + \beta_1\eta + \gamma_1\eta^2 + \cdots + \kappa_1\eta^{\nu-1}$$

is a polynomial $\chi(M)$ in M. We inquire if it will be possible to determine the parameters $\gamma, \delta, \cdots$, etc., in such a way that the set of conjugates will be cyclic; i. e., so that

$$M_{j+1} = \chi(M_j) \qquad (j = 0, 1, \cdots, \nu - 1).$$

Here we require that $M_\nu = M$. Since in this connection η behaves in all ways precisely as a scalar, save that $\eta^\nu = 0$, we can expand $\chi(M)$ in ascending powers of η as follows

$$\chi(M) = \chi(\alpha + \eta) = \chi(\alpha) + \eta\chi'(\alpha) + \tfrac{1}{2}\eta^2\chi''(\alpha) + \cdots + \frac{1}{(\nu - 1)!}\eta^{\nu-1}\chi^{(\nu-1)}(\alpha).$$

Since the matrices $1, \eta, \eta^2, \cdots, \eta^{\nu-1}$ are linearly independent the relation $M_1 = \chi(M)$ will be satisfied only if χ satisfies the conditions:

$$\chi(\alpha) = \alpha,\ \chi'(\alpha) = \beta_1,\ \chi''(\alpha) = 2\gamma_1, \cdots, \chi^{(\nu-1)}(\alpha) = (\nu - 1)!\,\kappa_1.$$

Expanding $\chi(M_j) = \chi(\alpha + \beta_j\eta + \gamma_j\eta^2 + \cdots + \kappa_j\eta^{\nu-1})$ in an entirely similar manner, we obtain:

$$\chi(M_j) = \chi(\alpha) + \eta\beta_j\chi'(\alpha) + \frac{\eta^2}{2!}\{\beta_j^2\chi''(\alpha) + 2\gamma_j\chi'(\alpha)\}$$
$$+ \frac{\eta^3}{3!}\{\beta_j^3\chi'''(\alpha) + 6\beta_j\gamma_j\chi''(\alpha) + 6\delta_j\chi'(\alpha)\} + \cdots.$$

In view of the above conditions on χ, it follows that $M_{j+1} = \chi(M_j)$ if, and only if, the parameters β, γ, δ, etc., are determined so as to satisfy the conditions: [6]

(23) $\beta_{j+1} = \beta_j \chi'(\alpha) = \beta_j \beta_1,$

(24) $\gamma_{j+1} = \frac{1}{2}\{\beta_j^2 \chi''(\alpha) + 2\gamma_j \chi'(\alpha)\} = \beta_j^2 \gamma_1 + \gamma_j \beta_1,$

(25) $\delta_{j+1} = \frac{1}{6}\{\beta_j^3 \chi'''(\alpha) + 6\beta_j \gamma_j \chi''(\alpha) + 6\delta_j \chi'(\alpha)\} = \beta_j^3 \delta_1 + 2\beta_j \gamma_j \gamma_1 + \delta_j \beta_1,$

(26) $\epsilon_{j+1} = \beta_j^4 \epsilon_1 + 3\beta_j^2 \gamma_j \delta_1 + (2\beta_j \delta_j + \gamma_j^2)\gamma_1 + \epsilon_j \beta_1,$

etc. Here we recall that $\gamma = \delta = \epsilon = 0$ and we agree to interpret $\gamma_\nu = \gamma$, $\delta_\nu = \delta$, etc.

Since $\beta_1 = \omega$, $\beta_j = \omega^j$, the first of these conditions is always satisfied. The remaining conditions are also obviously satisfied in the case in which the γ's, δ's, etc., are taken to be zero.

We therefore have the theorem:

THEOREM III. *If M is a matrix with reduced characteristic function $(\lambda - \alpha)^\nu$ so that M can be written in the form $\alpha + \eta$ where η is nilpotent of index ν, then the $\nu - 1$ matrices*

$$M_j = \alpha + \omega^j \eta \qquad (j = 1, \cdots, \nu - 1)$$

form a cyclic set of conjugates to M; that is, there exists a polynomial χ such that

$$M_{j+1} = \chi(M_j) \qquad (j = 0, 1, \cdots, \nu - 1;\ M_\nu = M).$$

We inquire, however, if there does not exist a more general cyclic set of conjugates than this.

Suppose for example that $\nu = 3$ so that $\beta_1 = \omega$ is a cube root of unity. We have then only to consider the conditions (24) on the γ's, which reduce in this case to the single condition

$$\gamma_1 + \gamma_2 = 0.$$

It follows then that the set of conjugates in (21) is always cyclic, whatever value be assigned to γ. This may also be verified by direct calculation.

For the special case $\nu = 3$ we can therefore state the following result: *If M is a matrix with reduced characteristic function $(\lambda - \alpha)^3$, every reduced*

---

[6] These conditions could have been derived more easily as follows. Since $\eta = M - \alpha$, we have on substituting into the expression for M_1,

$$M_1 = \chi(M) = \alpha + \beta_1(M - \alpha) + \gamma_1(M - \alpha)^2 + \cdots + \kappa_1(M - \alpha)^{\nu-1}.$$

If now we replace M by M_j, M_1 by M_{j+1} expand the expression on the right and equate coefficients of corresponding powers of η we obtain precisely the conditions (23), $\cdots$, (26). However, for later purposes, it is more instructive to proceed as we did above.

set of conjugates to M, which are expressible as polynomials in M, is a cyclic set.

Next consider the case $\nu = 4$. The conditions (24) yield in this case precisely the values of the γ's that are given in (22), while (25) reduces merely to the following

$$\delta_2 = 2\rho^2 i, \qquad \delta_1 + \delta_2 + \delta_3 = 0.$$

Hence the set of conjugates in (22) becomes cyclic provided only δ_2 be taken equal to $2\rho^2 i$. This also may be verified by direct calculation from (22).

Thus we find that in the case $\nu = 4$, although not every set of conjugates which are polynomials in M will constitute a cyclic set, the parameters γ, δ can be assigned values, not all zero, such that the resulting set will be cyclic.

In equations (8), $\cdots$, (12) we have not given general relations connecting the parameters $\beta, \gamma, \cdots, \kappa$ for $\nu > 5$. The derivation of these depends on the general formula for the n-th derivative of a function of a function which is admittedly quite complicated. We shall, therefore, not attempt to give a general proof by the present method, but shall outline the proof for $\nu = 5$. The method seems to be general.

Consider first equation (24) connecting the γ's:

$$(24) \qquad \gamma_{j+1} = \beta_1 \gamma_j + \gamma_1 \beta_j^2 \qquad (j = 0, 1, \cdots, 4).$$

Since $\gamma_0 = \gamma = 0$ by hypothesis, and since in order to complete the cycle we desire γ_5 to be equal to $\gamma = 0$, we have in (24) five homogeneous linear equations in the four unknowns $\gamma_1, \gamma_2, \gamma_3, \gamma_4$. The first of these reduces to the identity $\gamma_1 = \gamma_1$, so that there will be a non-zero solution for the four unknowns if, and only if, the four remaining equations are dependent. To show this dependence, multiply (24) through by β_1^{4-j} and sum as to j from 0 to 4. We get

$$\sum_0^4 \beta_1^{4-j}\gamma_{j+1} = \sum_0^4 \beta_1^{5-j}\gamma_j + \gamma_1 \sum \beta_1^{4+j}$$

since $\beta_j^2 = \beta_1^{2j}$. On the left change the index of summation by putting $j + 1 = i$. Since $\gamma_5 = \gamma_0 = 0$, this may be written:

$$\sum_0^4 \beta_1^{5-i}\gamma_i = \sum_0^4 \beta_1^{5-j}\gamma_j + \gamma_1\beta_1^4 \sum_0^4 \beta_1^{j}.$$

Since β_1 is a primitive fifth root of unity, the last term on the right is zero and the remaining terms cancel. The equations (24) are therefore dependent. Indeed, $\gamma_2, \gamma_3, \gamma_4$ are uniquely determined in terms of γ_1.

We need now show that these γ's satisfy also the necessary conditions (19). To do this multiply (24) through by β_j^k $(k = 0, 1, 2)$, and sum as to j. We get

$$\sum_{0}^{4} \beta_j{}^k \gamma_{j+1} = \beta_1 \sum \beta_j{}^k \gamma_j + \gamma_1 \sum \beta_j{}^{k+2}.$$

Since $\Sigma \beta_j{}^m = 0$ $(m = 2, 3, 4)$ the last summation on the right is zero. Also $\beta_j{}^k = \beta^k{}_{j+1}/\beta_1{}^k$, so that the above equation can be written, after changing as before the index of summation:

$$(1 - \beta_1{}^{k+1}) \sum \beta_j{}^k \gamma_j = 0 \qquad (k = 0, 1, 2).$$

Since $1 - \beta_1{}^{k+1} \neq 0$, we conclude that the γ's satisfying (24) satisfy (19) also.

In a similar manner we can show that there exist δ's and ϵ's not all zero which satisfy (25) and (26), and in addition satisfy (15) and (16). In fact by squaring (24) and summing it follows that for $\nu > 4$ the γ's satisfying (24) satisfy also the condition $\Sigma \gamma_j{}^2 = 0$, so that (16) reduces merely to $\Sigma \beta_i \delta_i = 0$.

As a very special solution we may take the γ's and δ's all zero and then we have only to determine the ϵ's satisfying (26). Thus, if we write $\beta_1 = \omega$, where ω is a primitive fifth root of unity, the conditions on the ϵ's become

$$\begin{aligned} \epsilon_2 &= (\omega + \omega^4)\epsilon_1 \\ \epsilon_3 &= \omega\epsilon_2 + \omega^3\epsilon_1 \\ \epsilon_4 &= \omega\epsilon_3 + \omega^2\epsilon_1 \\ 0 = \epsilon_5 &= \omega\epsilon_4 + \omega\epsilon_1. \end{aligned}$$

These equations yield the unique set of solutions

$$\epsilon_2 = (\omega + \omega^4)\epsilon_1, \qquad \epsilon_3 = -(\omega + \omega^4)\epsilon_1, \qquad \epsilon_4 = -\epsilon_1,$$

which obviously satisfy the additional condition $\Sigma \epsilon_j = 0$. Hence, in the case $\nu = 5$, we have the special set of conjugates:

$$\begin{aligned} M &= \alpha + \eta \\ M_1 &= \alpha + \omega\eta + \epsilon\eta^4 \\ M_2 &= \alpha + \omega^2\eta + (\omega + \omega^4)\epsilon\eta^4 \\ M_3 &= \alpha + \omega^3\eta - (\omega + \omega^4)\epsilon\eta^4 \\ M_4 &= \alpha + \omega^4\eta - \epsilon\eta^4. \end{aligned}$$

Here $M_{j+1} = \chi(M_j)$, where $\chi(\lambda) = \alpha + \omega(\lambda - \alpha) + \epsilon(\lambda - \alpha)^4$.

Moreover, in general, if $M = \alpha + \eta$, where η is nilpotent of index ν, and if we choose κ's such that

$$\begin{aligned} \kappa_0 &= 0, \\ \kappa_{j+1} &= \omega\kappa_j + \omega^{\nu-j}\kappa_1 \qquad (j = 0, 1, \cdots, \nu - 1) \end{aligned}$$

where ω is a primitive ν-th root of unity, it will follow that

$$\kappa_\nu = \kappa_0 = 0 \quad \text{and} \quad \sum_{1}^{\nu-1} \kappa_j = 0$$

and that the matrices

$$M_j = \alpha + \omega^j \eta + \kappa_j \eta^{\nu-1} \qquad (j = 1, \cdots, \nu - 1) \tag{27}$$

form a cyclic set of conjugates to M. Indeed, it is easy to verify that if $\chi(\lambda)$ is the polynomial

$$\chi(\lambda) = \alpha + \omega(\lambda - \alpha) + \kappa_1(\lambda - \alpha)^{\nu-1}$$

then

$$M_{j+1} = \chi(M_j) \qquad (j = 0, \cdots, \nu - 1).$$

We therefore have the theorem:

THEOREM IV. *If $M = \alpha + \eta$, where η is nilpotent of index ν, the matrices*

$$M_j = \alpha + \beta_j \eta + \gamma_j \eta^2 + \cdots + \kappa_j \eta^{\nu-1} \qquad (j = 1, \cdots, \nu - 1)$$

are polynomials in M, and the parameters $\beta, \gamma, \cdots, \kappa$ can always be so chosen, with the γ's, $\cdots$, κ's not all zero, such the M_j constitute a cyclic set of conjugates to M.

It will be observed that for $\kappa_1 = 0$, the set of conjugates (27) reduces to the set of Theorem III.

4. Extended sets of conjugates. Hitherto we have considered reduced sets of conjugates only, and we have shown that if $M = \alpha + \eta$, where η is nilpotent of index ν, we can determine all sets of matrices $M_1, \cdots, M_{\nu-1}$ which are polynomials in M and which possess the property that

$$\prod_{j=0}^{\nu-1} (\lambda - M_j) = (\lambda - \alpha)^\nu. \tag{6}$$

If now η is of index $\mu < \nu$, the argument by which we obtained the conjugates in the preceding section shows that precisely the same coefficients that were found in determining the M_j there will yield a set of $\nu - 1$ M's satisfying (6). With regard to the matrix M, this set will in this case be not a reduced set but an extended set of conjugates.

For example, if

$$M = \begin{pmatrix} \alpha & 1 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & \alpha \end{pmatrix} = \alpha + \eta \quad \text{where} \quad \eta = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

is nilpotent of index 2, the matrices in (21) with $\eta^2 = 0$

$$M = \alpha + \eta, \qquad M_1 = \alpha + \omega\eta, \qquad M_2 = \alpha + \omega^2\eta$$

are such that the last two constitute an extended set of conjugates to M. In this particular case, the reduced set of conjugates would consist of the single matrix $M_1 = \alpha - \eta$.

Before proceeding to a discussion of extended sets of conjugates we consider the following problem:

Let M be a matrix of the form $\alpha + \eta$, where η is nilpotent of index ν, and let m be a positive integer $< \nu$. Our problem is to find $m < \nu$ matrices N_j $(j = 1, \cdots, m)$ which are expressible as polynomials in M and which satisfy identically:

$$\prod_{j=1}^{m} (\lambda - N_j) = (\lambda - \alpha)^m. \tag{28}$$

If $m = 1$, N_1 must be a scalar α. We suppose then that $m > 1$. We may write N_1 in the form

$$N_1 = \alpha + a_1\eta + a_2\eta^2 + \cdots + a_{\nu-1}\eta^{\nu-1} = \alpha + \zeta,$$

where

$$\zeta = a_1\eta + a_2\eta^2 + \cdots + a_{\nu-1}\eta^{\nu-1} \tag{29}$$

is nilpotent of index $\leqq \nu$. If $a_1 \neq 0$, ζ is nilpotent of index ν, and by Theorem II we cannot find a set of N's including N_1 which satisfies (28).

Let a_k be the first coefficient in (29) that is different from zero. Then ζ is nilpotent of index μ where μ is such that

$$(\mu - 1)k < \nu \leqq \mu k.$$

We must determine k so that $\mu \leqq m$. Since $\nu > m$, we may write

$$\begin{aligned} \nu &= qm - r & (0 \leqq r < m) \\ \nu &= (q-1)m + r' & (0 < r' \leqq m), \end{aligned}$$

where q, r and r' are integers with $q > 1$. If we take $k \leqq q - 1$, then

$$km \leqq (q-1)m < \nu \leqq \mu k,$$

so that $\mu > m$. But if we take $k \geqq q$, then

$$km \geqq qm \geqq \nu > k(\mu - 1),$$

so that $m \geqq \mu$.

Let us then take $N_1 = \alpha + \zeta$, where

$$\zeta = a_q\eta^q + \cdots + a_{\nu-1}\eta^{\nu-1}.$$

ζ is then nilpotent of index $\mu \leqq m$ and we can determine, as in section 2, $\mu - 1$ matrices $N_2, \cdots, N_\mu$ which are polynomials in ζ and therefore in M, and such that

$$\prod_{j=1}^{\mu} (\lambda - N_j) = (\lambda - \alpha)^\mu.$$

If $\mu < m$, we may now adjoin to the set of μ matrices N_j $m - \mu$ matrices $N_{\mu+1}, \cdots, N_m$ which are scalars or polynomials in M, determined in the same manner as above, and such that for the entire set of N's we have

$$\prod_{j=1}^{m} (\lambda - N_j) = (\lambda - \alpha)^m.$$

It is clear that these matrices N can be built up in a great variety of ways.

Next suppose that $m > \nu$ in (28). Then an extended set of conjugates to M can be built up by the method mentioned at the beginning of the section. Otherwise, we may build up a reduced set of conjugates and adjoin to these $m - \nu$ scalars or matrices N built up in the manner just indicated.

Finally, we suppose that M is of the same form as before, *viz.*, $\alpha_1 + \eta$, and we set ourselves the problem of determining matrices M_j which are polynomials in M and which possess the property that

$$(30) \qquad (\lambda - M)(\lambda - M_1) \cdots (\lambda - M_{\tau-1}) = F(\lambda),$$

where

$$F(\lambda) = (\lambda - \alpha_1)^{\nu_1}(\lambda - \alpha_2)^{\nu_2} \cdots (\lambda - \alpha_r)^{\nu_r}, \qquad (\Sigma\nu_i = \tau).$$

Since $F(\lambda)$ must be divisible by the reduced characteristic function $(\lambda - \alpha_1)^{\nu}$ of M, this last condition is impossible if $\nu_1 < \nu$. We suppose then that $\nu_1 \geqq \nu$. We then determine first a set of conjugates $M_1, \cdots, M_{\nu_1 - 1}$, to M, which is either a reduced set or an extended set according as $\nu_1 = \nu$ or $\nu_1 > \nu$. These satisfy the condition:

$$\Pi(\lambda - M_j) = (\lambda - \alpha_1)^{\nu_1}.$$

Next we use the root α_2 and take $M_{\nu_1} = \alpha_2 + \zeta$ where ζ is of the form (29) and build up a set of $\nu_2 - 1$ conjugates to M_{ν_1}. These satisfy the condition

$$\Pi(\lambda - M_j) = (\lambda - \alpha_2)^{\nu_2}.$$

We proceed in the same way with regard to each of the factors $(\lambda - \alpha_i)^{\nu_i}$ of F. The set of matrices obtained by adjoining all these subsets are obviously commutative in pairs and satisfy (30).

5. The general square matrix M. We now turn to a consideration of a general square matrix M whose reduced characteristic function we shall suppose to be

$$(31) \qquad \phi(\lambda) = (\lambda - \alpha_1)^{\nu_1}(\lambda - \alpha_2)^{\nu_2} \cdots (\lambda - \alpha_r)^{\nu_r} \qquad (\Sigma\nu_i = \tau).$$

where the α's are distinct. In order to apply the results of the foregoing work to this case, we must first consider the principal idempotent and nilpotent elements of M.

6. The principal idempotent and nilpotent elements of M.[7] We sup-

[7] Wedderburn, "Lectures on matrices," *American Mathematical Society Publications*, 1934, pp. 28-29.

pose that the reduced characteristic function of M is given by (31) and that $r > 1$. Define the polynomials $h_i(\lambda)$ as follows:

$$h_i(\lambda) = \frac{\phi(\lambda)}{(\lambda - \alpha_i)^{\nu_i}} \qquad (i = 1, \cdots, r).$$

We can then determine two scalar polynomials $H_i(\lambda)$ and $G_i(\lambda)$ of degrees not exceeding $\nu_i - 1$ and $\tau - \nu_i - 1$, respectively, such that

$$H_i(\lambda) h_i(\lambda) + G_i(\lambda)(\lambda - \alpha_i)^{\nu_i} \equiv 1.$$

If we write

$$\phi_i(\lambda) = H_i(\lambda) h_i(\lambda) \qquad (i = 1, \cdots, r),$$

and for the matric polynomial $\phi_i(M)$ write merely ϕ_i, then ϕ_i is the principal idempotent element of M corresponding to the root α_i. These matrices ϕ_i satisfy the conditions

(32) $\qquad \phi_i^k = \phi_i \qquad$ for any positive integer k;

(33) $\qquad \phi_i \phi_j = 0 \qquad (i \neq j)$;

(34) $\qquad \sum_{i=1}^{r} \phi_i = I.$

Moreover, these ϕ's are linearly independent and none is zero.

Let us now denote by η_i the matric polynomial in M defined by

(35) $\qquad \eta_i = \eta_i(M) = (M - \alpha_i)\phi_i \qquad (i = 1, \cdots, r).$

It is easily shown that these matrices η_i satisfy the conditions

(36) $\qquad \eta_i^k \neq 0 \quad (k < \nu_i); \qquad \eta_i^{\nu_i} = 0,$

(37) $\qquad \eta_i \phi_i = \eta_i = \phi_i \eta_i; \qquad \eta_i \eta_j = 0 \quad (i \neq j),$

and moreover,

(38) $\qquad M = \Sigma(\alpha_i \phi_i + \eta_i) = \Sigma \phi_i (\alpha_i + \eta_i).$

The matrix η_i is the principal nilpotent element of M corresponding to the root α_i.

From the above relations it is easy to establish the following lemma:

LEMMA. *If $\psi(\lambda) = a + b\lambda + c\lambda^2 + \cdots + k\lambda^{\nu_i - 1}$ is any scalar polynomial in λ, the relation $\phi_i \psi(\eta_i) = 0$ is satisfied only if $\psi(\lambda) \equiv 0$, i. e., only if $a = b = \cdots = k = 0$.*

7. Sets of conjugates for the general matrix M. We set ourselves first the problem of determining $\tau - 1$ matrices M_j $(j = 1, \cdots, \tau - 1)$ which are expressible as polynomials in M and which in addition satisfy the condition:

(39) $(\lambda - M)(\lambda - M_1) \cdots (\lambda - M_{\tau-1}) \equiv \phi(\lambda) = (\lambda - \alpha_1)^{\nu_1} \cdots (\lambda - \alpha_r)^{\nu_r}$,

identically in λ. Such a set of M's automatically satisfies the condition I of section 1.

Since M can be written in the form (38) and since the ϕ's obey the relations (32) and (33), it follows readily that any polynomial $f(M)$ in M can be written

$$f(M) = \Sigma \phi_i f(\alpha_i + \eta_i) = \Sigma \phi_i f_i(\eta_i).$$

Conversely, since the ϕ's and η's are themselves polynomials in M, any expression of this last type, where the f_i are polynomials, is a polynomial in M. Let

$$M_j = \Sigma \phi_i f_i^{(j)}(\eta_i).$$

Since $\lambda = \lambda I = \Sigma \lambda \phi_i$, we may write (39) in the form

$$\Sigma \phi_i \prod_{j=0}^{\tau-1} \{\lambda - f_i^{(j)}(\eta_i)\} \equiv \Sigma \phi(\lambda) \phi_i,$$

which in view of (33) leads to

$$\phi_i \Pi \{\lambda - f_i^{(j)}(\eta_i)\} \equiv \phi(\lambda) \phi_i \qquad (i = 1, \cdots, r).$$

By the lemma of the preceding section, these conditions will hold if, and only if, the f's are so determined that

(40) $$\Pi \{\lambda - f_i^{(j)}(\eta_i)\} \equiv (\lambda - \alpha_1)^{\nu_1} \cdots (\lambda - \alpha_r)^{\nu_r} \qquad (i = 1, \cdots, r).$$

Consider first this last condition with $i = 1$. Since M can be written in the form (38), the first factor in the brackets on the left is $\lambda - (\alpha_1 + \eta_1)$, where η_1 is nilpotent of index ν_1. Our problem therefore reduces to finding a set $f_1^{(j)}(\eta_1)$ $(j = 1, \cdots, \tau - 1)$ which are polynomials in η_1 and which constitute an extended set of conjugates to $\alpha_1 + \eta_1$. This problem was solved completely in section 4. We arrange these in a column headed ϕ_1, the leading entry in the column being $\alpha_1 + \eta_1$.

We then proceed in the same way to determine a set of $\tau - 1$ conjugates $f_i^{(j)}(\eta_i)$ with respect to $\alpha_i + \eta_i$ for $i = 2, \cdots, r$ in turn. For each i, we arrange these in a column headed ϕ_i and with $\alpha_i + \eta_i$ as the leading entry. In this way we are led finally to an array of the type:

(41) $$\begin{array}{ccccc} \phi_1 & \cdots & \phi_i & \cdots & \phi_r \\ \alpha_1 + \eta_1 & \cdots & \alpha_i + \eta_i & \cdots & \alpha_r + \eta_r \\ f_1^{(1)}(\eta_1) & \cdots & f_i^{(1)}(\eta_i) & \cdots & f_r^{(1)}(\eta_r) \\ f_1^{(2)}(\eta_1) & \cdots & f_i^{(2)}(\eta_i) & \cdots & f_r^{(2)}(\eta_r) \\ \cdot & \cdots & \cdot & \cdots & \cdot \\ f_1^{(\tau-1)}(\eta_1) & \cdots & f_i^{(\tau-1)}(\eta_i) & \cdots & f_r^{(\tau-1)}(\eta_r). \end{array}$$

Now M is equal to the sum of the products of the ϕ's by the corresponding elements in the first row of the above array. In the same way, we may take as M_j $(j = 1, \cdots, \tau - 1)$ the sum of the products of the ϕ's by the corresponding elements in the $(j+1)$-st row. Clearly, in any column of the above any two rows after the first can be interchanged. Hence a set of $\tau - 1$ conjugates M_j can be built up in a great variety of ways.

As an illustration, let us consider a matrix M which has as its reduced characteristic function

$$\phi(\lambda) = (\lambda - 1)^3(\lambda + 1)^2(\lambda - 2).$$

Here $\alpha_1 = 1$, $\alpha_2 = -1$, $\alpha_3 = 2$, so that M can be written

$$M = \phi_1 + \eta_1 - \phi_2 + \eta_2 + 2\phi_3$$

where η_1 and η_2 are nilpotent of index 3 and 2, respectively. The array built up in the manner described above would be in this case:

$$(42) \qquad \begin{array}{ccc} \phi_1 & \phi_2 & \phi_3 \\ 1 + \eta_1 & -1 + \eta_2 & 2 \\ 1 + \omega\eta_1 + a\eta_1^2 & -1 - \eta_2 & 1 \\ 1 + \omega^2\eta_1 - a\eta_1^2 & 2 & 1 \\ -1 + b\eta_1^2 & 1 + c\eta_2 & 1 \\ -1 - b\eta_1^2 & 1 - c\eta_2 & -1 \\ 2 & 1 & -1 \end{array}$$

Using the elements in the successive rows as coefficients of the ϕ's, we would have as one set of conjugates to M the following:

$$M_1 = \phi_1(1 + \omega\eta_1 + a\eta_1^2) + \phi_2(-1 - n_2) + \phi_3;$$

$$\cdots\cdots\cdots\cdots$$

$$M_5 = 2\phi_1 \qquad\qquad + \phi_2 \qquad\qquad + \phi_3.$$

The set of conjugates given by Sokolnikoff for this particular case is obtained from the above on putting $a = b = c = 0$.

We now prove the theorem:

THEOREM V. *If for a matrix M there exists a completely cyclic set of conjugates, it is always possible to arrange the elements in the columns of the array* (41) *in such a way that each column will be cyclic.*

For let

$$M_j = \sum_{i=1}^{r} \phi_i f_i^{(j)}(\eta_i)$$
$$M_{j+1} = \sum \phi_i f_i^{(j+1)}(\eta_i),$$

and suppose that there exists a polynomial χ such that

$$M_{j+1} = \chi(M_j) \qquad (j = 1, \cdots, \tau - 1).$$

Since $\chi(M_j) = \Sigma\phi_i\chi[f_i^{(j)}(\eta_i)]$ this requires that

$$\Sigma\phi_i f_i^{(j+1)}(\eta_i) = \Sigma\phi_i\chi[f_i^{(j)}(\eta_i)] \qquad (j = 1, \cdots, \tau - 1),$$

and this, by (33) and the lemma of section 6, requires that

$$f_i^{(j+1)}(\eta_i) = \chi[f_i^{(j)}(\eta_i)] \qquad (i = 1, \cdots, r;\ j = 1, \cdots, \tau - 1).$$

The theorem is therefore proved.

Now a mere glance at the third column of the array (42) is sufficient to show that the elements in this column cannot be made cyclic by any rearrangement of the elements.

We therefore have the theorem:

THEOREM VI. *It is not possible to obtain a completely cyclic set of conjugates for the general matrix M.*

8. Special cases. Consider now the special case considered by Sokolnikoff in which the reduced characteristic function of M is of the type

$$\phi(\lambda) = (\lambda - \alpha_1)^\nu(\lambda - \alpha_2)^\nu \cdots (\lambda - \alpha_r)^\nu.$$

We may then write

$$M = \Sigma(\alpha_i\phi_i + \eta_i)$$

where each η_i is nilpotent of index ν. Let ω denote a primitive ν-th root of unity. Then as given by Sokolnikoff,

the first $\nu - 1$ conjugates are $\Sigma(\alpha_i\phi_i + \omega^j\eta_i)$ $\quad (j = 1, \cdots, \nu - 1)$,
the next ν conjugates are $\Sigma\alpha_{i+1}\phi_i$ $\quad (\alpha_{r+i} = \alpha_i)$
the next ν conjugates are $\Sigma\alpha_{i+2}\phi_i$

and so on, where the α's are cyclicly interchanged.

These may be generalized in two ways.

In the first place, instead of taking the second set of ν conjugates as above, we might take ν matrices exactly the same as $M, M_1, \cdots, M_{\nu-1}$ above in every respect save that α_i has been replaced by α_{i+1}, i. e., with the α's cyclicly interchanged. For the next ν conjugates we might take ν matrices of the same type with the α's cyclicly interchanged once more; and so on.

On the other hand, instead of taking as the first ν conjugates the matrices given above, we might take the following:

$$M_j = \Sigma\phi_i(\alpha_i + \beta_j\eta_i + \gamma_j\eta_i^2 + \cdots + \kappa_j\eta_i^{\nu-1}) \qquad (j = 1, \cdots, \nu - 1),$$

the β's, γ's, $\cdots$, κ's being so determined as to satisfy the conditions (23), $\cdots$, (26). Since $\beta_j = \omega^j$, these conjugates reduce to those mentioned earlier

in this section provided we put $\gamma_j = \delta_j = \cdots = \kappa_j = 0$. The next ν conjugates may then be taken as of the same form as the matrices $M, M_1, \cdots, M_{\nu-1}$ but with the α's cyclicly interchanged; and so on.

It is then easy to show that the matrices $M, M_1, \cdots, M_{\nu-1}$ constitute a cyclic subset among the conjugates to M. To prove this, we note that M_1 is known to be a polynomial $\chi(M)$ in M.[8] Then just as in section 3 it follows that χ must satisfy the conditions

$$(43)\quad \chi(\alpha_i) = \alpha_i,\ \chi'(\alpha_i) = \beta_1,\ \chi''(\alpha_i) = 2\gamma_1, \cdots, \chi^{(\nu-1)}(\alpha_i) = (\nu - 1)!\,\kappa_1.$$

In precisely the same way it can be shown that the further requirement $M_{j+1} = \chi(M_j)$ imposes only the additional conditions

$$\beta_{j+1} = \beta_j \chi'(\alpha_i) = \beta_j \beta_1,$$
$$\gamma_{j+1} = \tfrac{1}{2}[\beta_j{}^2 \chi''(\alpha_i) + 2\gamma_j \chi'(\alpha_i)] = \beta_j{}^2 \gamma_1 + \gamma_j \beta_1,$$

etc. But the β's, γ's, etc. were chosen in advance so as to satisfy these very conditions. Hence the statement is proved.

Similarly, it follows that each succeeding set of ν conjugates constitute a cyclic subset.

Consider now the r matrices of the above set:

$$M = \Sigma(\alpha_i \phi_i + \eta_i),$$
$$M_\nu = \Sigma(\alpha_{i+1} \phi_i + \eta_i),$$
$$\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot$$
$$M_{(r-1)\nu} = \Sigma(\alpha_{i+r-1} \phi_i + \eta_i),$$

where each is obtained from the preceding by a cyclic interchange on the α's. We know that M_ν is a polynomial in M. From the equation

$$M_\nu = \chi(M) = \Sigma \phi_i \chi(\alpha_i + \eta_i) = \Sigma \phi_i [\chi(\alpha_i) + \eta \chi'(\alpha_i) + \frac{n^2}{2} \chi''(\alpha_i) + \cdots]$$

it follows that

$$\chi(\alpha_i) = \alpha_{i+1},\ \chi'(\alpha_i) = 1,\ \chi''(\alpha_i) = \cdots = \chi^{(\nu-1)}(\alpha_i) = 0.$$

But this is precisely the condition that

$$M_{(j+1)\nu} = \chi(M_{j\nu}).$$

It is easy to see that for $\nu > 1$ the conjugates

$$M_1, M_2, \cdots, M_{r\nu-1}$$

cannot be chosen in such a way that the entire set, including M, will be cyclic.

[8] Since the degree of the reduced characteristic function of M is νr any polynomial in M can be expressed uniquely as an equivalent polynomial of degree at most $\nu r - 1$. If then χ be taken of degree $\leqq \nu r - 1$ the expression of $M_1 = \chi(M)$ is unique.

For if $M_1 = \chi(M)$, the polynomial χ is subject to the conditions (43), while the condition

$$M_\nu = \chi(M_{\nu-1})$$

would require that $\chi(\alpha_i) = \alpha_{i+1}$ which is incompatible with (43).

We therefore have the theorem:

THEOREM VII. *If the reduced characteristic function of a matrix M is of the form $\phi(\lambda) = (\lambda - \alpha_1)^\nu(\lambda - \alpha_2)^\nu \cdots (\lambda - \alpha_r)^\nu$, then for $\nu > 1$ there does not exist a set of conjugates M_j which are polynomials in M and such that the entire set, including M, will be cyclic. However, the conjugates can be so chosen that in the set there will exist cyclic subsets, such as*

$$(M, M_1, \cdots, M_{\nu-1}), \cdots, (M_{(r-1)\nu}, \cdots, M_{r\nu-1}),\ (M, M_\nu, \cdots, M_{(r-1)\nu}).$$

9. Matrices with linear elementary divisors. The case in which $\nu = 1$ is of particular interest. The roots of the minimum equation are then all distinct and M has linear elementary divisors. The conclusions here might be drawn from those in section 8 as corollaries.[9] However, it is believed that the case is of sufficient interest to warrant separate treatment.

Pierce[10] considered the case in which the characteristic equation of M has all distinct roots, which is precisely the situation considered originally by Taber for $n = 3$. Obviously, the situation considered by Pierce is a special case of the one treated in this section.

If $\nu = 1$ each η_i is zero so that M can be written

$$M = \Sigma \alpha_i \phi_i \tag{44}$$

and for any polynomial $\chi(M)$ we have

$$\chi(M) = \Sigma \chi(\alpha_i) \cdot \phi_i. \tag{45}$$

The array (41) has in this case as elements in its first row the numbers $\alpha_1, \alpha_2, \cdots, \alpha_r$. The columns are then built up by filling in with the α's in such a way that in each column each α is used once and once only. Here each of the conjugates M_j will be a polynomial in M, but it will not be true in general either that each M_j has the same reduced characteristic function as M or that the set will be cyclic.

For example, if $r = 3$ the array (41) might be

ϕ_1	ϕ_2	ϕ_3
α_1	α_2	α_3
α_2	α_1	α_1
α_3	α_3	α_2

[9] Sokolnikoff, *loc. cit.*, p. 175.

[10] Pierce, *Bulletin of the American Mathematical Society*, vol. 36 (1936), pp. 262-264.

corresponding to which we should have the set of conjugates

$$
\begin{aligned}
M &= \alpha_1\phi_1 + \alpha_2\phi_2 + \alpha_3\phi_3 \\
M_1 &= \alpha_2\phi_1 + \alpha_1\phi_2 + \alpha_1\phi_3 \\
M_2 &= \alpha_3\phi_1 + \alpha_3\phi_2 + \alpha_2\phi_3.
\end{aligned}
$$

The reduced characteristic function of M_1 is $(\lambda - \alpha_2)(\lambda - \alpha_1)$, while that of M_2 is $(\lambda - \alpha_3)(\lambda - \alpha_2)$. Moreover, if $M_1 = \chi(M)$ it follows from (45) that

$$\chi(\alpha_1) = \alpha_2, \qquad \chi(\alpha_2) = \alpha_1, \qquad \chi(\alpha_3) = \alpha_1.$$

But then $M_2 \neq \chi(M_1)$.

It is possible, however, to arrange the α's in the columns of (41) in such a way that both of these conditions will be satisfied. Indeed all we have to do is to arrange them in such a way that no α occurs more than once in any row or column. One way to do this is to interchange the α's cyclicly. Thus, if M is given in (44), then we might take as M_j the matrices

$$M_j = \Sigma \alpha_{i+j}\phi_i \qquad (j = 1, \cdots, r-1).$$

where for $i + j$ we take the least positive residue (mod r). Obviously if $M_1 = \chi(M)$, so that $\chi(\alpha_i) = \alpha_{i+1}$, then also

$$M_{j+1} = \chi(M_j).$$

Moreover, each M_j built up in this way has the same *reduced* characteristic function as M. However, it should be noted that this is not necessarily true as to the *characteristic* function.

THEOREM VIII. *If a matrix M has linear elementary divisors so that it can be written in the form* (44), *then for $r > 2$ there always exists more than one set of conjugates to M which are expressible as polynomials in M. It will not be true in general that each of the conjugates will have the same reduced characteristic function as M or that the set of conjugates will be cyclic. It is always possible, however, to find a set of conjugates satisfying both of these conditions. One such set is that obtained from the expression* (44) *of M by repeated cyclic interchanges on the α's.*

If $r = n$, the characteristic function and the reduced characteristic function are identical. In this case we have the

COROLLARY. *If the characteristic roots $\alpha_1, \cdots, \alpha_n$ of a matrix M are distinct there exists a polynomial $\chi(\lambda)$ (which is unique if its degree be taken $\leqq n-1$) such that*

$$\alpha_{i+1} = \chi(\alpha_i) \qquad (i = 1, \cdots, n;\ \alpha_{n+1} = \alpha_1).$$

If then we define

$$M_1 = \chi(M), M_2 = \chi(M_1), \cdots, M_{n-1} = \chi(M_{n-2}),$$

the matrices M_j *all have the same characteristic function as* M *and constitute a cyclic set of conjugates to* M.

10. Williamson's matrix.[11] In particular, if the characteristic function of M is of the form

$$\lambda^n - k^n \qquad (k \neq 0) \tag{46}$$

so that the characteristic roots of M are

$$\alpha_1 = k, \alpha_2 = \omega k, \cdots, \alpha_n = \omega^{n-1}k,$$

where ω is a primitive n-th root of unity, then the unique polynomial $\chi(\lambda)$ of degree $\leqq n-1$ such that $\chi(\alpha_i) = \alpha_{i+1}$ is precisely $\chi(\lambda) = \omega$. This is the case considered by Williamson.

Williamson's matrix is the product $B = \Omega A$ of the diagonal matrix $\Omega = [1, \omega, \cdots, \omega^{n-1}]$ by the circulant

$$A = \begin{pmatrix} a_0 & a_1 & \cdots & a_{n-1} \\ a_{n-1} & a_0 & \cdots & a_{n-2} \\ \cdot & \cdot & \cdot & \cdot \\ a_1 & a_2 & \cdots & a_0 \end{pmatrix}.$$

If we can show that the characteristic function of B is of the form (46) with $k \neq 0$, then the matrices

$$\omega B, \omega^2 B, \cdots, \omega^{n-1} B \tag{47}$$

constitute a set of conjugates to B.

Write $\omega_i = \omega^i$, so that

$$1, \omega_1, \omega_2, \cdots, \omega_{n-1}$$

are the n-th roots of unity. The matrix

$$P = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & \omega_1 & \cdots & \omega_{n-1} \\ 1 & \omega_1^2 & \cdots & \omega^2_{n-1} \\ \cdot & \cdot & \cdot & \cdot \\ 1 & \omega_1^{n-1} & \cdots & \omega^{n-1}_{n-1} \end{pmatrix}$$

is non-singular and, as is well known,[12] if we denote by $\theta(\omega_i)$ the polynomial

[11] Williamson, *American Mathematical Monthly*, vol. 39 (1932), pp. 280-285. See also Udo Wegner, "The product of a circulant matrix and a special diagonal matrix," *American Mathematical Monthly*, vol. 40 (1933), pp. 23-25.

[12] Scott, *Theory of Determinants*, Cambridge (1880), p. 82.

$$\theta(\omega_i) = a_0 + a_1\omega_i + a_2\omega_i^2 + \cdots + a_{n-1}\omega_i^{n-1},$$

we have

$$AP = PN, \quad \text{or} \quad P^{-1}AP = N$$

where N is the diagonal matrix $[\theta(1), \theta(\omega_1), \cdots, \theta(\omega_{n-1})]$. Moreover, it is easy to verify that $\Omega P = PR$, where R is the matrix

$$R = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

Hence, $P^{-1}BP = P^{-1}\Omega PP^{-1}AP = RN$, where

$$RN = \begin{pmatrix} 0 & 0 & \cdots & 0 & \theta(\omega_{n-1}) \\ \theta(1) & 0 & \cdots & 0 & 0 \\ 0 & \theta(\omega_1) & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & \theta(\omega_{n-2}) & 0 \end{pmatrix}.$$

The characteristic function of this last matrix, and therefore of $B = \Omega A$, is then easily seen to be

$$\lambda^n - \theta(1)\theta(\omega_1)\cdots\theta(\omega_{n-1}) = \lambda^n - k^n$$

where k^n is equal to the determinant of ΩA or of $(-1)^{n-1}A$. If then $|A| \neq 0$ it follows that the matrices (47) form a cyclic set of conjugates to B.

But if $|A| = 0$, the characteristic function of B reduces to λ^n so that B is nilpotent of index $\nu \leqq n$. It follows then that

$$(\lambda - B)(\lambda - \omega B)\cdots(\lambda - \omega^{n-1}B) \equiv \lambda^n - B^n \equiv \lambda^n$$

so that the matrices (47) constitute a set, although not necessarily a *reduced* set, of conjugates to B.

11. Sets of conjugates for a derogatory matrix. In the case in which M has a single elementary divisor $(\lambda - \alpha_i)^{\nu_i}$ corresponding to each root α_i of the reduced characteristic equation $\phi(\lambda) = 0$ in (31), then the latter coincides with the characteristic equation. M is then said to be *non-derogatory,* and the only matrices M_j commutative with M are polynomials in M. In this case, therefore, all sets of conjugates to M are found by the methods of the preceding sections.

If, however, for at least one root α_i of $\phi(\lambda) = 0$ there is more than one

elementary divisor, the foregoing statement is no longer true. We proceed now to an investigation of this case.

Suppose for example that, corresponding to the root α_1, M has the elementary divisors

$$(\lambda-\alpha_1)^{\nu_1}, (\lambda-\alpha_1)^{\nu'_1}, \cdots, (\lambda-\alpha_1)^{\nu_1^{(t)}}. \tag{48}$$

Then it is known that corresponding to the set of elementary divisors (48) there exists a set of so-called *partial* idempotent elements

$$\psi_0, \psi_1, \cdots, \psi_t,$$

and a set of partial nilpotent elements

$$\zeta_0, \zeta_1, \cdots, \zeta_t$$

the latter being nilpotent of indices

$$\nu_1, \nu', \cdots, \nu_1^{(t)}$$

respectively.

The ψ's satisfy relations similar to those satisfied by the principal idempotent elements ϕ_i in (32), (33) and (34); *viz.*,

$$\psi_i^k = \psi_i; \qquad \psi_i\psi_j = 0 \quad (i \neq j) \qquad \sum_{i=0}^{t} \psi_i = \phi_1,$$

where ϕ_1 is the principal idempotent element corresponding to the root α_1. These ζ's satisfy relations similar to (35), (36) and (37) satisfied by the principal nilpotent elements, *viz.*,

$$\zeta_i^k \neq 0 \quad (k < \nu_1^{(i)}), \qquad \zeta_i^{\nu_1^{(i)}} = 0, \qquad \zeta_i\psi_i = \zeta_i = \psi_i\zeta_i,$$
$$\zeta_i\zeta_j = 0 \quad (i \neq j); \qquad \Sigma\zeta_i = \eta_1$$

where η_1 is the principal nilpotent element corresponding to the root α_1.

These partial elements are not in general polynomials in M. Moreover they are not unique. In fact, if to the matrix M we apply the transformation $T^{-1}MT$, where T is commutative with M, a particular set of partial elements is in general transformed into another set, and all such sets are obtained in this way.

It should be clear from the properties of the partial elements that any matrix expressible as a polynomial in a set of partial elements of M is commutative with M, and further that any two matrices both of which are expressible as polynomials in the same set of partial elements are commutative with each other.

We seek then conjugates M_j $(j = 1, \cdots, r-1)$ which are expressible

as polynomials in an arbitrary set of partial elements of M. It will then follow, just as in the discussion of section 7, that we are led to the problem of finding polynomials f satisfying equations of the form (40), where now η_1 has been replaced by each of the ζ_i in turn. In other words, in the array (41) the column headed ϕ_1 is split up into $t+1$ columns headed $\psi_0, \psi_1, \cdots, \psi_t$, respectively, and with the same root α_1 in all the columns.

In the first of these subcolumns, headed ψ_0, the leading entry is $\alpha_1 + \zeta_0$, where ζ_0 is nilpotent of index ν_1. The remaining $\tau - 1$ places in the first column are then to be filled in with matrices which are expressible as polynomials in ζ_0, and which constitute an extended set of conjugates to $\alpha_1 + \zeta_0$. These matrices satisfy a relation of the type (40) with η_1 replaced by ζ_0. In the second of these sub-columns headed ψ_1 the leading entry is $\alpha_1 + \zeta_1$. The remaining entries are matrices constituting an extended set of conjugates to $\alpha_1 + \zeta_1$, and satisfying a relation of the type (40) with η_1 replaced by ζ_1. We proceed in this way through all the $t+1$ columns, using $\zeta_2, \zeta_3, \cdots, \zeta_t$ in turn.

If there is more than one elementary divisor corresponding to the root α_2, we split up the column headed ϕ_2 into partial columns and proceed in the same way. Finally, we obtain an array of the type (41) but containing just as many columns as M has elementary divisors.

As the conjugate M_j $(j = 1, \cdots, \tau - 1)$ we may then take the sum of the products of the ψ's by the corresponding elements in the $(j+1)$-st row.

For a different set of partial elements clearly we are led to a different set of conjugates. We obtain in this way all sets of conjugates that are expressible as polynomials in the partial elements of M.

Consider for example a matrix of order 8 whose elementary divisors are

$$(\lambda - 1)^3, \qquad (\lambda - 1)^2, \qquad (\lambda + 1)^2, \qquad (\lambda - 2).$$

The partial elements corresponding to the last two elementary divisors are actually principal elements, so that the last two columns in the array (42) remain the same as there. However, the first column splits into two, as follows:

$$\begin{array}{cccc}
\psi_0 & \psi_1 & \phi_2 & \phi_3 \\
1 + \zeta_0 & 1 + \zeta_1 & -1 + \eta_2 & 2 \\
1 + \omega\zeta_0 + a\zeta_0^2 & 1 - \zeta_1 & -1 - \eta_2 & 1 \\
1 + \omega^2\zeta_0 - a\zeta_0^2 & 1 & 2 & 1 \\
-1 + b\zeta_0^2 & -1 + d\zeta_1 & 1 + c\eta_2 & 1 \\
-1 - b\zeta_0^2 & -1 - d\zeta_1 & 1 - c\eta_2 & -1 \\
2 & 2 & 1 & -1.
\end{array}$$

If then we write

$$
\begin{aligned}
&M = (1+\zeta_0)\psi_0 + (1+\zeta_1)\psi_1 + (-1+\eta_2)\phi_2 + 2\phi_3,\\
\text{then}\quad &M_1 = (1+\omega\zeta_0 + a\zeta_0^2)\psi_0 + (1-\zeta_1)\psi_1 + (-1-\eta_2)\phi_2 + \phi_3.\\
&M_2 = (1+\omega^2\zeta_0 - a\zeta_0^2)\psi_0 + \psi_1 + 2\phi_2 + \phi_3,
\end{aligned}
$$

etc.

Thus in the case where M is derogatory we have shown how to obtain many reduced sets of conjugates to M, but they are such that the M_j are not expressible as polynomials in M. However, even this method does not yield all sets of conjugates to M, since there exist matrices commutative with M which are not expressible as polynomials in any set of partial elements of M.

For example if

$$
M = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad M_1 = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad M_2 = \begin{pmatrix} 1 & -2 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix};
$$

it is easy to verify that the M's are commutative in pairs and moreover satisfy the condition I of section 1, with $\phi(\lambda)$ replaced by $(\lambda-1)^3$, the characteristic function of M. Hence, M_1 and M_2 constitute an extended set of conjugates to M, but neither is expressible as a polynomial in any set of partial elements of M. Such sets of conjugates we leave out of consideration in the present paper, since the methods employed do not lend themselves readily to a discussion of them. They are found most readily, perhaps, by a method given by P. Franklin in his paper "Algebraic matric equations," *Journal of Mathematics and Physics*, vol. 10 (1932), pp. 289-314.

THE UNIVERSITY OF NORTH CAROLINA.

ABELIAN FIELDS AND DUALITY OF ABELIAN GROUPS.*

By Reinhold Baer.

It is a classical theorem that the cyclic extensions of a (commutative) field are exactly those which are generated by solving an equation $x^n - k = 0$, provided sufficiently many roots of unity are available. Generalizing this result, Witt[1] has recently characterized the finite abelian extensions by means of invariants. It is the aim of this note to show that Witt's theory may be extended—almost without any modification—to all those extensions whose group is abelian (finite or not) and that this may be done by the simple device of substituting Pontrjagin's duality theory[2] for the classical theory of characters of finite abelian groups.

Chapter I. Vector Groups and Their Characters.

Instead of introducing a topology in the considered Galois groups and character groups it seems to be more appropriate to describe their elements as vectors whose coördinates are elements in certain finite groups and are subject to some conditions. This chapter is devoted to an account of these concepts and in particular to the theory of duality between abelian groups without elements of infinite order and abelian vector groups. Definitions and theorems are adapted to the needs of the applications in the second chapter and topological ideas have been avoided.

1. The concept of vector group. Two objects are needed for the definition of a vector group: a set S of groups and an operation $h(X < Y)$ defined for certain pairs X, Y in S. The groups in S will be called the coördinate groups and the group Y in S will be said to cover X in S, whenever $h(X < Y)$ exists. S and h will be subject to certain conditions which will be enunciated in the course of this section and which will be denoted by (1. A) and so on.

* Presented to the American Mathematical Society, September 6, 1937. Received by the Editors, June 16, 1937.

[1] E. Witt, "Der Existenzsatz für abelsche Funktionenkörper," *Journal für die reine und angewandte Mathematik*, Bd. 173 (1935); E. Witt, "Zyklische Körper und Algebren der Characteristik p vom Grad p^n," *Journal für die reine und angewandte Mathematik*, Bd. 176 (1937), pp. 126-140.

[2] L. Pontrjagin, "Theory of topological commutative groups," *Annals of Mathematics*, vol. 35 (1934), pp. 361-388; E. R. van Kampen, "Locally bicompact abelian groups and their character groups," *Annals of Mathematics*, vol. 36 (1935), pp. 448-463.

(1. A) *$h(X < X)$ exists for every X in S and is the identity automorphism of the group X.*
(1. B) *If Y covers X, then $h(X < Y)$ is a homomorphism, mapping the whole group Y upon the whole group X.*
(1. C) *If Z covers Y and Y covers X, then Z covers X and*

$$h(X < Z) = h(Y < Z)h(X < Y).$$

(1. D) *To any pair of groups in S there exists a group in S which covers both.*
(1. E) *To every group X in S and every normal subgroup N of X there exists one and only one group W in S such that X covers W and the homomorphism $h(W < X)$ maps exactly the elements in N upon 1.*
(1. F) *Every group in S is finite.*

It is a consequence of (1. A) and (1. E) that $h(X < Y)$ is an isomorphism if, and only if, $X = Y$ *and h is the identity.*

LEMMA 1. 1. *Suppose that S and h satisfy* (1. A) *to* (1. F).

(a) *If the group H in S covers the groups K and L in S, and if every element in H which is mapped upon* 1 *by $h(K < H)$ is also mapped upon* 1 *by $h(L < H)$, then K covers L.*
(b) *If F is a finite, not vacuous subset of S, then there exists one and only one group G in S such that G covers every group X in F and such that* 1 *is the only element in G which is mapped upon* 1 *by every homomorphism $h(X < G)$ for X in F. If furthermore H in S covers every group in F, then H covers G and G may therefore be called the least common cover* (= *l. c. c.*) *of the groups in F.*
(c) *If R is a not vacuous subset of S, then there exists one and only one group E in S which is covered by every group X in F and which covers every group in S which is covered by every group in F. E may therefore be called the greatest common ground* (= *g. c. g.*) *of the groups in R.*

Proof. Suppose that H covers the groups K and L. Denote by K' those elements in H *which are mapped upon* 1 *by* $h(K < H)$ and by L' those elements in H which are mapped upon 1 by $h(L < H)$. Assume that $K' \leq L'$. Then $h(K < H)$ maps L' upon a normal subgroup L'' of K and there exists therefore by (1. E) a group M in S such that K covers M and $h(M < K)$ maps exactly the elements in L'' upon 1. It follows now from (1. C) that H covers M and that $h(M < H)$ maps exactly the elements in L' upon 1 and it follows from (1. E) that $L = M$, i. e. that K covers L.

Suppose now that F is a finite, not vacuous subset of S. Then it follows

by complete induction from (1. C) and (1. D) that there exists at least one group H in S which covers every group X in F. Denote by H' the set of all those elements in the common cover H of F which are mapped upon 1 by every $h(X < H)$ for X in F. There exists by (1. E) a group G in S such that H covers G and $h(G < H)$ maps exactly the elements in H' upon 1, since H' is by (1. B) a normal subgroup of H. Now it follows from (a) that G is a common cover of the groups in F and it follows from the choice of G that 1 is the only element in G which is mapped upon 1 by every $h(X < G)$ for X in F.

Suppose now that $G(i)$ is—for $i = 1, 2$—a group in S such that $G(i)$ covers every X in F and such that 1 is the only element in $G(i)$ which is mapped upon 1 by every $h(X < G(i))$ for X in F. As has been proved just now, there exists a group G in S such that G covers $G(1)$ and $G(2)$ and such that 1 is the only element in G which is mapped upon 1 by both $h(G(1) < G)$ and $h(G(2) < G)$. It follows from (1. C) that G covers every X in S and that $h(X < G) = h(G(i) < G)h(X < G(i))$ for X in F and $i = 1, 2$. Thus if w is an element of G which is mapped upon 1 by $h(G(j) < G)$ for $j = 1$ or 2, then w is mapped upon 1 by every $h(X < G)$ for X in F and consequently w is mapped upon 1 by $h(G(j \pm 1) < G)$ and this implies that $w = 1$. Thus it follows from (1. E) and (1. A) that $G(1) = G = G(2)$.

It follows now from the proof that every common cover of the groups in F covers their uniquely determined l. c. c.

Denote by T the set of all those groups in S which are covered by every group in the subset R of S. It follows from (1. E) and (1. F) that T is finite, and T is not vacuous, since every group in S covers by (1. E) the uniquely determined group in S which consists of the identity only. It is now an obvious consequence of (b) that the l. c. c. of T is the g. c. g. of F and this completes the proof of the lemma.

If S is any set of groups, then it is usual to define as a vector any single valued function v which maps every group X in S on a uniquely determined element v_X of X, its X-coördinate. Vectors are multiplied in multiplying their coördinates and thus the vectors over S form a group. But if in the set S a covering operation $h(X < Y)$ is defined which satisfies (1. A) to (1. F), then only a certain subgroup will be considered, namely the subgroup of all those vectors v whose coördinates v_X satisfy:

(1. V) *If X covers Y, then $v_Y = v_X^{h(Y<X)}$.*

The group of all the vectors of S which satisfy (1. V) will be called the *vector group $V(S, h)$, defined over S by means of the covering operation $h(X < Y)$.*

If in particular S is a finite set, then it follows from Lemma 1. 1, (b) that $V(S, h)$ is essentially the l. c. c. of S, and that S is essentially the set of all the different factor groups of $V(S, h)$.

If the set S is infinite, but countable, then there exists a sequence $H(i)$ $(i = 1, 2, \cdots)$ of groups in S such that $H(i) \neq H(j)$ for $i \neq j$, $H(i)$ covers $H(j)$ for $j < i$ and such that every group X in S is covered by some $H(i)$. The vector group $V(S, h)$ consists essentially of the sequences $h(i)$ with $h(i)$ in $H(i)$ and $h(i) = h(i+1)^{h(H(i) < H(i+1))}$. The sequences which satisfy $h(i) = 1$ form a normal subgroup $V(i)$ of $V(S, h)$ and $V/V(i)$ and $H(i)$ are isomorphic. These groups are therefore essentially the so-called Cantorian groups.[3]

In the following the set S need not be countable and two further assumptions will be needed which are a consequence of (1. A) to (1. F), if S is at most countable.

(1. G) *If x is an element in the group X in S, then there exists a vector in $V(S, h)$ whose X-coördinate is x.*

In mapping every vector of $V(S, h)$ upon its X-coördinate a homomorphism of $V(S, h)$ upon the whole group X is defined, provided (1. G) is satisfied. If $V(X)$ is the set of all vectors in $V(S, h)$ whose X-coördinate is 1, then $V(X)$ is a normal subgroup and X and $V(S, h)/V(X)$ are essentially the same. Furthermore S is essentially equal to the set of these factor groups and h to the homomorphism induced by $V(S, h)$.

(1. H) *The subset T of the set S is the whole set S if, and only if, T satisfies the following conditions:*

(i) *If X is contained in T, then every group, covered by X, is contained in T.*

(ii) *If X and Y are contained in T, then T contains a group which covers both X and Y.*

(iii) *The vector v in $V(S, h)$ is the 1-vector if, and only if, $v_X = 1$ for every X in T.*

2. Subgroups of vector groups. It will be assumed in this section 2. that the set S of coördinate groups, the covering operation $h(X < Y)$ and the vector group $V(S, h)$ satisfy (1. A) to (1. H).

Notations. If U is a subgroup of $V(S, h)$ and X a group in S, then U_X is the set of all the X-coördinates of vectors in U, $S(U)$ is the set of all those

[3] D. van Dantzig, "Zur topologischen Algebra III," *Comp. Math.*, Bd. 3 (1936), pp. 408-426; H. Freudenthal, "Entwicklungen von Räumen und ihren Gruppen," *Comp. Math.* Bd. 4 (1937), pp. 145-234.

groups X in S which satisfy $U_X = 1$; $V(U)$ denotes the set of all the vectors v in $V(S,h)$ such that $v_X = 1$ for every X in $S(U)$ and $\bar{U}$ denotes the set of all the vectors v in $V(S,h)$ such that v_X is for every X in S an element of U_X.

$$V(U) \text{ and } \bar{U} \text{ are subgroups of } V(S,h).$$
$$S(V(U)) = S(U), \bar{U}_X = U_X \text{ for every } X \text{ in } S.$$
$$V(V(U)) = V(U), \bar{\bar{U}} = \bar{U}.$$

If U is a normal subgroup of $V(S,h)$, then it follows from (1. G) that every U_X is a normal subgroup of X, and if conversely every U_X is a normal subgroup of X, then $\bar{U}$ is a normal subgroup of $V(S,h)$.

(2. 1) $U \leq \bar{U} \leq V(U)$ *for every subgroup* U *of* $V(S,h)$.

This is a consequence of the identity: $S(U) = S(\bar{U}) = S(V(U))$.

(2. 2) *If* U *is a normal subgroup of* $V(S,h)$, *then* $V(U) = \bar{U}$.

Proof. There exists by (1. E) to every group X in S a uniquely determined group X' in S such that X covers X' and $h(X' < X)$ maps exactly the elements in U_X upon 1. If X in S covers Y in S, then it follows from (1. V) that U_X is mapped by $h(Y < X)$ exactly upon U_Y. This implies in particular that $U_{X'} = 1$ for every X in S, i. e. that every X' is an element of $S(U)$, and $S(U)$ is therefore by (1. A) exactly the set of all the groups X' for X in S. If v is a vector in $V(U)$, then $v_{X'} = 1$ for every X in S and now it follows from the choice of the group X' that v_X is an element of U_X for every X in S, i. e. that v is a vector in $\bar{U}$. Hence $V(U) \leq \bar{U}$ for normal subgroups U of $V(S,h)$ and consequently $V(U) = \bar{U}$ by (2. 1).

DEFINITION 2. 3. *The subgroup* U *of* $V(S,h)$ *is closed, if* $U = V(U)$.

Thus all subgroups $V(U)$ are closed subgroups and normal subgroups U are closed, if $U = \bar{U}$.

If G is any group whatsoever, then *denote by* $C(G)$ *its commutator group* and by G^m for a positive integer m the subset of those elements of G which mod $C(G)$ are m-th powers of elements of G. Then $C(G)$ and G^m are characteristic subgroups of G, $G/C(G)$ and G/G^m are abelian groups and the orders of the elements of G/G^m are finite and divisors of m.

(2. 4) $C(V(S,h))_X = C(X)$; $(V(S,h)^m)_X = X^m$ *for every* X in S.

3. Characters of groups without elements of infinite order.

Notation. E is the essentially uniquely determined group with the following properties:

(i) *Every finite subset of E is contained in a cyclic subgroup of E;*
(ii) *E contains elements of every finite order.*

This group E is abelian, contains for every positive integer n exactly one cyclic subgroup of order n and is isomorphic with the multiplicative group of the roots of unity in the field of complex numbers.

Suppose now that G is an abelian group without elements of infinite order. Then every homomorphism of G into a subgroup of E is called a *character of* G and the product fg of the characters f and g of G is defined by the equation: $x^{(fg)} = x^f x^g$ for every x in G. Thus the characters of G form an abelian group.

Since homomorphisms map elements of the finite order n upon elements whose order is a divisor of n, as a rule not the whole group E will be needed for the definition of the characters and it is obvious which part of E is really necessary.

Every character of G induces a character in every subgroup of G and conversely it may be proved in the usual way[4] that every character of every subgroup of G is induced by a character of the whole group G. It is a consequence of these facts, of the finiteness of the orders of the elements in G and of the condition (ii), that 1 *is the only element in G which is mapped upon* 1 *by every character of G.*

If U is any subgroup of G, then denote by $H(U)$ the group of all the characters of U. Let $S = S(G)$ be the set of all the groups $H(U)$ for finite subgroups U of G. If $U \leq U'$, then every character of U' induces a character of U and thus a homomorphism $h(H(U) < H(U'))$ of $H(U')$ upon the whole group $H(U)$ is defined. S and h satisfy obviously (1. A) to (1. C) and (1. F). (1. D) is a consequence of the fact that the join of two finite subgroups of G is a finite subgroup of G and (1. E) is a theorem in the classical theory of characters of finite abelian groups.[5]

If c is any character of G, then c induces in every finite subgroup U of G a character $c_{H(U)}$, its $H(U)$-coördinate. The characters of G are completely determined by its $H(U)$-coördinates, since G is the join of its finite subgroups. By the definition of the homomorphisms $h(H(U) < H(U'))$ it follows that the coördinates of the characters of G satisfy (1. V) and thus it may be proved that *the group of characters of G is exactly the vector group $V(S, h)$, defined over $S(G)$ by means of h.*

(1. G) is a consequence of the fact that every character of a subgroup of G is induced by a character of G and (1. H) follows from the remark: if U is any subgroup of G and w an element of G which is not contained in U, then

[4] Cp. footnote [2].

[5] E. Hecke, *Vorlesungen über die Theorie der algebraischen Zahlen*, 1923.

there exists a character of G which maps every element of U upon 1, but w upon an element $\neq 1$.

If the group G is not abelian, but all the elements of $G/C(G)$ are of finite order, then every character of G ($=$ homomorphism of G into E) maps the elements of $C(G)$ upon 1 and induces therefore a character of $G/C(G)$. Conversely every character of $G/C(G)$ is induced by exactly one character of G. Thus the characters of G are essentially the same as the characters of $G/C(G)$ and it is obvious how to apply the above considerations upon the theory of characters of groups G such that the elements of $G/C(G)$ are of finite order.

4. Characters of vector groups. If G is any group, f a homomorphism of G, then denote by $(G;f)$ the set of all those elements in G which are mapped upon 1 by f. $(G;f)$ is a normal subgroup of G and f defines an isomorphism of $G/(G;f)$ upon G^f.

In this section it shall be assumed that the set S of groups, the covering operation $h(X < Y)$ and the vector group $V(S,h)$ defined over S by means of h satisfy the conditions (1. A) to (1. H).

LEMMA 4. 1. *Let f be a homomorphism of $V(S,h)$ into the group E, defined in 3. Then there exists a group N in S such that $v^f = 1$, whenever $v_N = 1$ if, and only if, $(V(S,h);f)$ is a closed subgroup of the vector group $V(S,h)$ and $V(S,h)^f$ is a finite (and therefore a cyclic) subgroup of E.*

Proof. Assume first that there exists a group N in S such that $v_N = 1$ implies $v^f = 1$. Then $v_N = w_N$ implies $v^f = w^f$ and it follows therefore from (1. G) that f induces a homomorphism of N into E. $N^f = V(S,h)^f$ and (1. F) imply now that $V(S,h)^f$ is a finite and therefore a cyclic subgroup of E. A vector v belongs to $(V(S,h);f)$ if, and only if, its N-coördinate v_N belongs to $(N;f)$. Let now X be any group in S. Then there exists by (1. D) a common cover X' of X and N in S. If $v_{X'} = 1$, then $v_N = 1$ and therefore $v^f = 1$. Thus f is also a homomorphism of X' and the vector v belongs to $(V(S,h);f)$ if, and only if, its X'-coördinate belongs to $(X';f)$. But since $(X';f) = (V(S,h);f)_{X'}$ and since X' covers X, it follows that v belongs to $(V(S,f);f)$ if, and only if, its X-coördinate belongs to $(V(S,h);f)_X$. Thus $(V(S,h);f) = \overline{(V(S,h);f)}$ and $(V(S,h);f)$ is therefore, as a normal subgroup of $V(S,h)$, by (2. 2) a closed subgroup of $V(S,h)$.

Assume now that $(V(S,h);f)$ is a closed subgroup of $V(S,h)$ and that $V(S,h)^f$ is a finite and therefore a cyclic subgroup of E. There exists a vector g in $V(S,h)$ such that g^f generates the finite cyclic group $V(S,h)^f$. If $g_X{}^i$ is for every X in S an element of $(V(S,h);f)_X$, then g^i is an element

of $(V(S,h);f)$ since $(V(S,h);f)$ is a closed subgroup of the vector group $V(S,h)$. If n is the order of g mod $(V(S,h);f)$ and $0 < i < n$, then there exists therefore a group $N(i)$ in S such that $g^i{}_{N(i)}$ is not an element of $(V(S,h);f)_{N(i)}$. By Lemma 1.1 there exists the l. c. c. N of these groups $N(i)$ in S. From the choice of the groups $N(i)$ and N it follows that g^i is an element of $(V(S,h);f)$ if, and only if, $g_N{}^i$ is an element of $(V(S,h);f)_N$. Let now v be any vector in $V(S,h)$. Then there exists an integer i such that vg^i is an element of $(V(S,h);f)$. Consequently $(vg^i)_N = v_N g_N{}^i$ is an element of $(V(S,h);f)_N$. Now $v_N = 1$ implies that $g_N{}^i$ is an element of $(V(S,h);f)_N$ and therefore that g^i is an element of $(V(S,h);f)$. Thus $v_N = 1$ implies that v itself is an element of $(V(S,h);f)$ and this completes the proof of the Lemma.

Definition 4.2. *A character of the vector group $V(S,h)$ is a homomorphism f of $V(S,h)$ into E such that $(V(S,h);f)$ is closed and $V(S,h)^f$ is finite.*

Corollary 4.3. *To every character f of the vector group $V(S,h)$ there exists one and only one group N_f in S such that $v^f = 1$ if, and only if, $v_{N_f} = 1$. N_f is uniquely determined by either of the following two properties:*

f induces an isomorphism in N_f.

The group X in S covers N_f if, and only if, $v_X = 1$ implies $v^f = 1$.

This Corollary follows essentially from Lemma 4.1 and (1.E).

By means of this Corollary it is rather simple to construct all the possible characters of the vector group $V(S,h)$. Just choose any cyclic group Z in S and an isomorphism f of Z into E. Z and f determine uniquely a character f of $V(S,h)$ by the equation $v^f = v_z{}^f$, and $Z = N_f$. Conversely every character of $V(S,h)$ leads to such a pair Z, f which is uniquely determined by the character. This implies in particular that the vector v is mapped upon 1 by every character of the vector group $V(S,h)$ if, and only if, all its coördinates are elements of the respective commutator groups. *Thus if $V(S,h)$ is abelian, then 1 is the only vector mapped upon 1 by every character of the vector group $V(S,h)$.*

5. The duality theorem. Suppose that the set S of groups, the covering operation $h(X < Y)$ in S and the vector group $V(S,h)$, defined over S by means of h, satisfy the conditions (1.A) to (1.H), that G is a group such that all the elements of $G/C(G)$ *are of finite order*, and that an operation $g \circ v$ is defined which obeys the following rules:

(5. A) *$g \circ v$ is for every g in G and for every v in $V(S,h)$ a uniquely determined element of the group E, defined in section 3.*

(5. B) $(gh) \circ v = (g \circ v)(h \circ v)$, $g \circ (vw) = (g \circ v)(g \circ w)$.

(5. C) *The group $(V(S,h); g)$ of the elements v in $V(S,h)$ which satisfy $g \circ v = 1$ is for every g in G a closed subgroup of the vector group $V(S,h)$.*

These rules imply that every v in $V(S,h)$ induces by $g \circ v$ a character of the group G, and consequently $c \circ v = 1$ for every c in $C(G)$. Thus every g in G induces a homomorphism of $V(S,h)$ into a finite subgroup of E and it follows now from (5. C) that every element g in G induces by $g \circ v$ a character of the vector group $V(S,h)$.

It is of importance to note that for the considerations of this section not the whole group E is needed, but only a subgroup of E which contains elements of order n for every positive integer n which is the order of an element in $G/C(G)$ or the order of an element in some group $X/C(X)$ for X in S.

THEOREM 5. 1. *The following four properties of the group G, the vector group $V(S,h)$ and the operation $g \circ v$ are equivalent:*

(a) *$g \circ v = 1$ for every g in G if, and only if, $v = 1$ and*

$g \circ v = 1$ for every v in $V(S,h)$ if, and only if, $g = 1$.

(b) *$g \circ v = 1$ for every g in G if, and only if, $v = 1$,*

G is abelian and every character of G is induced by some v in $V(S,h)$.

(c) *$V(S,h)$ is abelian, every character of the vector group $V(S,h)$ is induced by some element in G and*

$g \circ v = 1$ for every v in $V(S,h)$ if, and only if, $g = 1$.

(d) *G and $V(S,h)$ are each the character group of the other group (and therefore abelian).*

Remark. If G and (or) $V(S,h)$ are *finite* groups, then the above Theorem 5. 1 is a well known fact in the theory of characters of finite abelian groups [5] and the propositions (a) to (d) together with the finiteness-assumption imply that G and $V(S,h)$ are *isomorphic*. This remark shall be used in the course of the proof of the Theorem 5. 1.

Proof. A. If G is abelian and every character of G is induced by an element of $V(S,h)$, then

$$g \circ v = 1 \text{ for every } v \text{ if, and only if, } g = 1,$$

as follows from a remark in section **3**.

If $V(S, h)$ is abelian and every character of the vector group $V(S, h)$ is induced by some element in G, then

$$g \circ v = 1 \text{ for every } g \text{ if, and only if, } v = 1,$$

as follows from a remark at the end of section **4**.

These facts imply that (b) and (c) are both consequences of (d) and that (a) is a consequence of each of the propositions (b) and (c). Thus the Theorem 5.1 will be proved as soon as it has been shown that (d) is a consequence of (a).

B. Assume that (a) is satisfied.

Clearly both G and $V(S, h)$ are abelian groups, since the elements in G and in $V(S, h)$ induce homomorphisms into an abelian group.

If g is any element of the group G, then there exists by (5.C) and Corollary 4.3 a uniquely determined group N_g in S such that $g \circ v = 1$ if, and only if, $v_{N_g} = 1$. If F is any finite subgroup of G, then there exists by Lemma 1.1 the l. c. c. N_F of the groups N_g for g in F. Since N_F covers each N_g with g in F, and since 1 is the only element of N_F which is mapped upon 1 by every $h(N_g < N_F)$ with g in F, it follows from the definition of N_g that $g \circ v = 1$ for every g in F if, and only if, $v_{N_F} = 1$, and N_F is uniquely determined by this property. Since therefore

$$g \circ v = g \circ w \text{ for every } g \text{ in } F \text{ if, and only if, } v_{N_F} = w_{N_F},$$

an operation $g \circ x$ is defined for the elements g in F and x in N_F by the equations

$$g \circ x = g \circ v \text{ for every } v \text{ in } V(S, h) \text{ such that } v_{N_F} = x$$

and this operation between F and N_F satisfies (5.A) and (5.B). From (a) and the definition of $g \circ x$ it follows that $g \circ x = 1$ for every x implies $g = 1$, and it follows from the definition of N_F and the definition of $g \circ x$ that $g \circ x = 1$ for every g in F implies $x = 1$. It is therefore a consequence of the theory of characters of finite abelian groups—as has been mentioned in the Remark—that F and N_F are isomorphic finite abelian groups and each the character group of the other group (under the operation $g \circ x$).

The fundamental step in the proof will be the proof of the following statement:

(5. 2) *N_F defines a one-one-correspondence between the set of all the finite subgroups F of G and the whole set S.*

Suppose first that F and F' are two finite subgroups of G such that $N_F = N_{F'} = M$. Denote by F'' the subgroup of G which is generated by the elements in F and in F'. F'' is a finite subgroup of G, since G is abelian. Since $g \circ v = 1$ for every g in F if, and only if, $v_M = 1$, and since $g' \circ v = 1$ for every g' in F' if, and only if, $v_M = 1$, it follows that $g'' \circ v = 1$ for every g'' in F'' if, and only if, $v_M = 1$. Thus $N_{F''} = M$ and the groups F, F', F'' and M are isomorphic. Since F and F' are subgroups of the isomorphic finite group F'', it follows that $F = F'' = F'$ and N_F defines therefore a one-one-correspondence between the set of all finite subgroups F of G and a certain subset T of S.

It is an obvious consequence of the definition of N_F that N_F covers $N_{F'}$ if, and only if, $F' \leq F$. If F and F' are any two finite subgroups of G, F'' their join-group, then $N_{F''}$ is a common cover of N_F and $N_{F'}$. If U is any subgroup of a group N_F, then denote by F' the group of all those elements g in F which satisfy $g \circ u = 1$ for every u in U. Since F and N_F are each the character group of the other group, it follows from the classical theory of characters of finite abelian groups [5] that $N_{F'}$ is covered by N_F and that exactly the elements in U are mapped upon 1 by $h(N_{F'} < N_F)$.—If finally v is a vector such that $v_{N_F} = 1$ for every finite subgroup F of G, then $f \circ v = 1$ for every element f of G and it follows from (a) that $v = 1$. The set T satisfies therefore the conditions (i) to (iii) of (1. H) and this implies $T = S$, thus completing the proof of (5. 2).

If $\boldsymbol{c}$ is any character of the vector group $V(S, h)$, then there exists by Corollary 4. 3 a group N_c in S such that $v^c = 1$ if, and only if, $v_{N_c} = 1$. The character $\boldsymbol{c}$ of $V(S, h)$ induces therefore a character $\boldsymbol{c}$ of the cyclic group N_c. There exists by (5. 2) a uniquely determined finite subgroup F of G such that $N_F = N_c$ and F and N_F are each the character group of the other group. F contains therefore an element g such that $g \circ x = x^c$ for every element x in N_c and consequently $g \circ v = v^c$ for every v in $V(S, h)$. Thus every character of the vector group $V(S, h)$ is induced by an element of G.

If $\boldsymbol{c}$ is any character of the group G, then $\boldsymbol{c}$ induces in every finite subgroup F of G a character $\boldsymbol{c}_F$. Since N_F is the character group of F, there exists an element c_F in N_F such that $g \circ c_F = g^{c_F}$ for every g in F. These elements c_F satisfy (1. V) and they are therefore the coördinates of a vector v in $V(S, h)$. If (g) is the cyclic subgroup of G, generated by g, then $g \circ v = g \circ v_{N_{(g)}} = g^{c_{(g)}} = g^c$ and the character $\boldsymbol{c}$ of G is therefore induced by the element v of $V(S, h)$. This completes the proof of the Theorem 5. 1.

If the group G, the vector group $V(S,h)$ and the operation $g \circ v$ do not satisfy the properties (a) to (d) of the Theorem 5.1, then at least the following statements may be derived from this theorem:

denote by G_0 the set of all those elements g in G such that $g \circ v = 1$ for every v in $V(S,h)$ and denote by V_0 the set of all vectors v, satisfying $g \circ v = 1$ for every g in G; then

$$C(G) \leq G_0, \qquad \overline{C(V(S,h))} \leq V_0.$$

The operation $X \circ Y$, defined by $G_0 g \circ V_0 v = g \circ v$ for the elements X of G/G_0 and the elements Y of $V(S,h)/V_0$ satisfies (5. A) to (5. C) and the conditions (a) to (d) of Theorem 5.1. Every character of G is induced by an element of $V(S,h)$ and every character of the vector group $V(S,h)$ is induced by an element of the group G if, and only if, $G_0 = C(G)$ and $V_0 = \overline{C(V(S,h))}$.

Chapter II. Characters of Galois Groups and Characterization of Abelian Extensions.

6. Representation of Galois groups as vector groups. This section is devoted to a representation of Galois groups as vector groups and to an exposition of those parts of Galois theory which are needed for this purpose.

Notations. If F is a field and $\boldsymbol{G}$ a group of automorphisms of F, then $J(\boldsymbol{G})$ is the subfield of all those elements f in F which satisfy $f = f^{\boldsymbol{g}}$ for every $\boldsymbol{g}$ in $\boldsymbol{G}$.—If K is a subfield of F, then an automorphism $\boldsymbol{g}$ of F is called a K-automorphism of F, if $s = s^{\boldsymbol{g}}$ for every s in K, and the group of all the K-automorphisms of F is denoted by $\boldsymbol{R}(K < F)$.

THEOREM 6.1. *$\boldsymbol{G} = \boldsymbol{R}(J(\boldsymbol{G}) < F)$ for every finite group $\boldsymbol{G}$ of automorphisms of F.*

A Proof of this theorem may be added, since this theorem is slightly more general than the theorems which are usually reproduced in the literature, and since the proof itself is simpler than the proofs of the classical case use to be.[6]

It is obvious that $\boldsymbol{G} \leq \boldsymbol{R}(J(\boldsymbol{G}) < F)$. Assume now that the theorem is wrong. Then there exists an automorphism $\boldsymbol{w}$ in $\boldsymbol{R}(J(\boldsymbol{G}) < F)$ which is not contained in $\boldsymbol{G}$. There exists therefore to every element $\boldsymbol{g}$ in $\boldsymbol{G}$ an element $v(\boldsymbol{g})$ in F such that $v(\boldsymbol{g})^{\boldsymbol{g}} \neq v(\boldsymbol{g})^{\boldsymbol{w}}$.

Let now correspond to every element $\boldsymbol{g}$ in $\boldsymbol{G}$ a new symbol $x(\boldsymbol{g})$. If H is any field, then denote by $H^* = H(\cdots x(\boldsymbol{g}) \cdots)$ the field of all rational func-

[6] The idea of the proof goes back to a paper of O. Bolza.

tions in the $x(g)$ with coefficients in H, and if h is an automorphism of H, then denote by h^* the uniquely determined automorphism of H^* which coincides in H with h and has the $x(g)$ as fixed elements.

Let G^* be the group of the automorphisms g^* of F^* for g in G. Then $J(G^*) = J(G)^*$. Since G is a finite group, it is possible to introduce the quantity $t = \sum_{g \text{ in } G} v(g)x(g)$ in F^* and the polynomial

$$f(x) = f(x; \cdots x(g) \cdots) = \prod_{h \text{ in } G} (x - t^{h^*}).$$

Since the application of an automorphism g^*, for g in G, upon $f(x)$ effects a permutation of the factors, defining $f(x)$, the coefficients of the polynomial $f(x)$ are elements in $J(G^*)$ and the coefficients of the polynomial $f(x; \cdots x(g) \cdots)$ are elements of $J(G)$. Since w belongs to $R(J(G) < F)$, it follows that w^* belongs to $R(J(G^*) < F^*)$ and consequently that t^{w^*} is a solution of $f(x) = 0$. Hence there exists an automorphism z in G such that $t^{z^*} = t^{w^*}$. But since $t^{h^*} = \sum_{g \text{ in } G} v(g)^h x(g)$, this implies that $v(z)^z = v(z)^w$ in contradiction to the choice of the elements $v(g)$ and this completes the proof of the theorem.

THEOREM 6.2. *Let U be a subfield of the field F. Then every element of F is mapped by U-automorphisms of F only upon a finite number of different elements in F and $U = J(R(U < F))$ if, and only if, every element in F is a solution of an equation in U whose degree is exactly the number of its different solutions in F (i. e. if F is an algebraic, normal and separable extension of U).*

This may be proved by a slight change in the usual procedure.[7] The decisive step may be stated as a lemma for future reference.

LEMMA 6.3. *If the field F is an algebraic and normal extension of its subfield U, then every U-isomorphism of a field between U and F upon a field between U and F is induced by an automorphism of F.*[8]

Suppose now that the field F is an algebraic, normal and separable extension of its subfield K. A field B between K and F is *normal* with regard to K, if every K-automorphism of F maps the field B upon B, and B is *finite*

[7] R. Baer, "Abbildungseigenschaften algebraischer Erweiterungen," *Mathematische Zeitschrift*, Bd. 33 (1931), pp. 451-479.

[8] E. Steinitz, *Algebraische Theorie der Körper*. Neu herausgegeben und mit einem Anhang: Abriss der Galoisschen Theorie versehen von Reinhold Baer und Helmut Hasse, Berlin, 1930. Cp. Lemma 1, Lemma 1a and Lemma 2 on pp. 134-136.

with regard to K, if B may be generated by adjoining a finite number of elements to K.

Denote by $S(K < F)$ the set of all the fields B between K and F which are finite and normal with regard to K and denote by $\boldsymbol{S}(K < F)$ the set of all groups $\boldsymbol{R}(K < B)$ for B in $S(K < F)$. If B and B' are two fields in $S(K < F)$ and $B \leq B'$, then every K-automorphism of B' induces a K-automorphism of B and every K-automorphism of B is induced, by Lemma 6.3, by some automorphism of B'. Thus a homomorphism $\boldsymbol{h}(\boldsymbol{R}(K < B) < \boldsymbol{R}(K < B'))$ of $\boldsymbol{R}(K < B')$ upon the whole group $\boldsymbol{R}(K < B)$ is defined. The set $\boldsymbol{S}(K < F)$ and the covering operation $\boldsymbol{h}$ satisfy obviously (1. A) to (1. C) and (1. F). That they satisfy (1. D) is a consequence of the fact that the field B'', generated by the fields B and B' in $S(K < F)$, belongs to $S(K < F)$. If B is a field in $S(K < F)$ and $\boldsymbol{N}$ a normal subgroup of $\boldsymbol{R}(K < B)$, then $J(\boldsymbol{N})$ is a field between K and B which is normal and finite with regard to K. $\boldsymbol{N}$ is finite, since $\boldsymbol{R}(K < B)$ is finite. Hence $\boldsymbol{N} = \boldsymbol{R}(J(\boldsymbol{N}) < F)$ by Theorem 6.1 and $\boldsymbol{h}(\boldsymbol{R}(K < J(\boldsymbol{N}))) < \boldsymbol{R}(K < B))$ maps exactly the elements in $\boldsymbol{N}$ upon 1. Thus it has been shown that (1. A) *to* (1. F) *are fulfilled by the set* $\boldsymbol{S}(K < F)$ *and the covering operation* $\boldsymbol{h}$.

Every K-automorphism $\boldsymbol{g}$ of F induces an automorphism $\boldsymbol{g}_B$ in every field B in $S(K < F)$. $\boldsymbol{g}_B$ belongs to $\boldsymbol{B} = \boldsymbol{R}(K < B)$ and these coördinates satisfy (1. V). $\boldsymbol{g}$ is uniquely determined by the $\boldsymbol{g}_B$, since F is the join of the fields B in $S(K < B)$. If in every group $\boldsymbol{B}$ in $\boldsymbol{S}(K < F)$ an automorphism $\boldsymbol{g}_B$ has been chosen in such a way that (1. V) is satisfied, then these $\boldsymbol{g}_B$ determine exactly one K-automorphism of F whose coördinates they are and this proves that *the group* $\boldsymbol{R}(K < F)$ *of the K-automorphisms of F is exactly the vector group, defined over the set* $\boldsymbol{S}(K < F)$ *by means of the covering operation* $\boldsymbol{h}$.

(1. G) is satisfied in this vector group, since by Lemma 6.3 every K-automorphism of a field B in $S(K < F)$ is induced by an automorphism of F. Suppose now that $\boldsymbol{T}$ is a subset of the set $\boldsymbol{S}(K < F)$ which satisfies the conditions (i) to (iii) of (1. H). Denote by T the set of the corresponding fields in $S(K < F)$ and by W the join of the fields in T. Then W is a field between K and F such that the identity is the only W-automorphism of F. It follows therefore from Theorem 6.2 that $W = F$ and this implies that $T = S(K < F)$ and consequently that $\boldsymbol{T} = \boldsymbol{S}(K < F)$, i. e. that also (1. H) is satisfied. The result of these considerations may be stated as a theorem.

THEOREM 6.4. *If the field F is an algebraic, normal and separable extension of its subfield K, then the set* $\boldsymbol{S}(K < F)$, *the covering operation* $\boldsymbol{h}$ *and the vector group* $\boldsymbol{V}(\boldsymbol{S}, \boldsymbol{h})$, *defined over* $\boldsymbol{S}(K < F)$ *by means of* $\boldsymbol{h}$, *obey the postulates* (1. A) *to* (1. H) *and* $\boldsymbol{V}(\boldsymbol{S}(K < F), \boldsymbol{h}) = \boldsymbol{R}(K < F)$.

COROLLARY 6.5. *If* $\boldsymbol{G}$ *is a subgroup of the vector group* $\boldsymbol{R}(K < F)$, *then* $\boldsymbol{R}(J(\boldsymbol{G}) < F) = \boldsymbol{V}(\boldsymbol{G})$*—in the notation of section 2—and if* $\boldsymbol{G}$ *is a normal subgroup, then—by* (2.2)*—*$\boldsymbol{R}(J(\boldsymbol{G}) < F) = \bar{\boldsymbol{G}}$.

7. Realization of the character group of the Galois group. Let F be an algebraic, normal and separable extension of the field K. It will be possible to use the group $E(K)$ of the roots of unity in K for the definition of the characters of the vector group $\boldsymbol{R}(K < F)$ if, and only if, the following hypothesis is satisfied:

(7.R) *K contains n different n-th roots of unity, if there are elements of order n in a group* $\boldsymbol{B}/\boldsymbol{C}(\boldsymbol{B})$ *for* $\boldsymbol{B}$ *in the set* $\boldsymbol{S}(K < F)$.

NOTATION. K^* is the multiplicative group of all the elements $\neq 0$ in K and $G^*(F)$ *is the multiplicative group of all those elements* $\neq 0$ *in F whose order mod K^* is a positive integer n such that K contains n different n-th roots of unity.*

THEOREM 7.1. *Let F be an algebraic, normal and separable extension of the field K and G^* a group between K^* and $G^*(F)$.*

(a) *The operation $X \circ \mathbf{g}$, defined for the elements X of the group G^*/K^* and the elements $\mathbf{g}$ of the vector group $\mathbf{R}(K < F)$ by the equation*

$$K^*u \circ \mathbf{g} = u^{1-\mathbf{g}}$$

obeys the postulates (5.A) *to* (5.C).

(b) *The following properties of the field F, the group G^* and the operation $X \circ \mathbf{g}$ are equivalent:*

(1) *F is the field, generated by G^*.*

(2) $\boldsymbol{R}(K < F)$ *is abelian, F satisfies* (7.R) *and* $G^* = G^*(F)$.

(3) G^*/K^* *and* $\mathbf{R}(K < F)$ *are—by means of the operation $X \circ \mathbf{g}$—each the character group of the other group.*

Proof. The definition of $X \circ \boldsymbol{v}$ is independent of the particular choice of the representative of the class X, since the elements $\boldsymbol{v}$ are K-automorphisms of F and consequently $k^{1-\boldsymbol{v}} = 1$ for k in K^*. If the order of the element K^*g of G^*/K^* is n, then g is a solution of the equation $x^n - g^n = 0$ with coefficients in K. $K^*g \circ \boldsymbol{v}$ is therefore for every K-automorphism $\boldsymbol{v}$ of F an

n-th root of unity and is by the choice of G^* an element of K, i. e. an element of $E(K)$.

$$(XY) \circ \boldsymbol{v} = (X \circ \boldsymbol{v})(Y \circ \boldsymbol{v}),$$

since $\boldsymbol{v}$ is an automorphism of F, and

$$X \circ (\boldsymbol{vw}) = X^{1-vw} = X^{1-w}X^{w-vw} = (X \circ \boldsymbol{w})(X \circ \boldsymbol{v})^{w} = (X \circ \boldsymbol{w})(X \circ \boldsymbol{v})$$

since $X \circ \boldsymbol{v}$ is an element of K and $\boldsymbol{w}$ a K-automorphism. Thus (5. A) and (5. B) are satisfied. That (5. C) is satisfied, is a consequence of Lemma 4. 1 and the fact that every X in G^*/K^* is contained in a field B in $S(K < F)$.

If $X \circ \boldsymbol{v} = 1$ for every $\boldsymbol{v}$ in $\boldsymbol{R}(K < F)$, then X belongs to $J(\boldsymbol{R}(K < F))$ and since this field equals K by Theorem 6. 2 it follows that $X = K^* = 1$, i. e.

(7. 2) $\qquad X \circ \boldsymbol{v} = 1$ for every $\boldsymbol{v}$ if, and only if, $X = 1$.

If (1) is satisfied and $\boldsymbol{v}$ is an element of $\boldsymbol{R}(K < F)$ such that $X \circ \boldsymbol{v} = 1$ for every X in G^*/K^*, then $f^{v} = f$ for every f in F, i. e. $\boldsymbol{v} = \mathbf{1}$ and it follows from (7. 2) that (a) of Theorem 5. 1 is satisfied. Theorem 5. 1 may be applied, as follows from Theorem 7. 1, (a), and the choice of G^* and consequently (d) of Theorem 5. 1 holds true. Hence (3) is a consequence of (1).

If (3) is satisfied, denote by W the field, generated by G^*. This is a field between K and F. Let $\boldsymbol{v}$ be any W-automorphism of F. Then $X \circ \boldsymbol{v} = 1$ for every X in G^*/K^*, since these X are subsets of W. Since (3) holds true, also (d) of Theorem 5. 1 holds true and therefore (a) of Theorem 5. 1 is satisfied. Thus $\boldsymbol{v} = \mathbf{1}$ and it follows from Theorem 6. 2 that $W = F$, i. e. (1) is a consequence of (3).

Since (a) of Theorem 7. 1 has been proved for every group G^* between K^* and $G^*(F)$, it may be applied upon $G^*(F)$ itself. The same applies to (7. 2). Thus if (3) holds true, there exists to every X in $G^*(F)/K^*$ an X' in G^*/K^* such that $X \circ \boldsymbol{v} = X' \circ \boldsymbol{v}$ for every $\boldsymbol{v}$ in $\boldsymbol{R}(K < F)$, and since this implies $X = X'$, it follows that (3) implies the equality of G^* and $G^*(F)$, i. e. (2).

Assume now that c is a character of the vector group $\boldsymbol{R}(K < F)$. Then there exists by Corollary 4. 3 a uniquely determined group $\boldsymbol{N}_c$ in $\boldsymbol{S}(K < F)$ such that $\boldsymbol{v}^c = 1$ if, and only if, $\boldsymbol{v}_{\boldsymbol{N}_c} = \mathbf{1}$. Denote by B_c the field in $S(K < F)$ such that $\boldsymbol{N}_c = \boldsymbol{R}(K < B_c)$. Then c is a character of the group $\boldsymbol{N}_c$. There exists in B_c an element u such that $b = \sum_{h \text{ in } N_c} \boldsymbol{h}^c u^h \neq 0$—this follows e. g. from the existence of a normal basis [9] of B_c with regard to K—and, using the fact

[9] M. Deuring, "Galoissche Theorie und Darstellungstheorie," *Mathematische Annalen*, Bd. 107 (1932), pp. 140-144.

that by (7. R) all the elements $\boldsymbol{h}^c$ are contained in K, it may be verified that $b^{1-h} = \boldsymbol{h}^c$. If n is the order of $\boldsymbol{N}_c$, then $b^n = \prod_{h \text{ in } N_C} b^h \boldsymbol{h}^c = b^{\Sigma h} \prod \boldsymbol{h}^c$ is an element of K and K^*b is consequently an element of $G^*(F)/K^*$. Since $K^*b \circ \boldsymbol{v} = \boldsymbol{v}^c$ for every $\boldsymbol{v}$ in $\boldsymbol{R}(K < F)$, it follows that (3) is a consequence of (2) and this completes the proof of the Theorem.

COROLLARY 7. 3. *Let F be an algebraic, normal and separable extension of the field K, satisfying* (7. R).

(i) *$\boldsymbol{R}(K < F)$ is abelian if, and only if, F is the field, generated by $G^*(F)$.—If $\boldsymbol{R}(K < F)$ is abelian, then $G^*(F)/K^*$ and the vector group $\boldsymbol{R}(K < F)$ are each the character group of the other group.*

(ii) *$J(\boldsymbol{C}(\boldsymbol{R}(K < F)))$ is the field, generated by $G^*(F)$.*

(iii) *Denote by $G^*_m(F)$ the group of elements in $G^*(F)$ whose order mod K^* is a divisor of m and by F_m′the field, generated by $G^*_m(F)$, then $G^*(F_m) = G^*_m(F)$ and $F_m = J(\boldsymbol{R}(K < F)^m)$.*

(i) is an obvious consequence of the Theorem 7. 1 and (ii) and (iii) follow in using Corollary 6. 5 and (2. 4).

8. Characteristic invariants of abelian extensions.

DEFINITION 8. 1. *$e(K < F)$ is for every algebraic, normal and separable extension F of the field K the set of all the orders of elements in groups $\boldsymbol{B}/\boldsymbol{C}(\boldsymbol{B})$ for $\boldsymbol{B}$ in $\boldsymbol{S}(K < F)$ and*

$P(n, K < F)$ is for every positive integer n the group of all the elements $\neq 0$ in K which are n-th powers of elements in F.

$P(n, K < F)$ contains the group K^{*n} of the n-th powers of elements $\neq 0$ in K. If K contains exactly n different n-th roots of unity, then $P(n, K < F) = G^*_n(F)^n$ and *the correspondence between the elements of $G^*_n(F)$ and their n-th powers induces an isomorphism between $G^*_n(F)/K^*$ and $P(n, K < F)/K^{*n}$.*

THEOREM 8. 2. *Suppose that H and L are algebraic, normal and separable extensions of the field K, that $\boldsymbol{R}(K < H)$ and $\boldsymbol{R}(K < L)$ are abelian and that* (7. R) *is satisfied by H and L. There exists a K-isomorphism between H and L if, and only if,*

(a) $\quad e(K < H) = e(K < L) = e;$

(b) $\quad P(n, K < H) = P(n, K < L) = P(n)$ *for every n in e.*

Proof. It is obvious that (a) and (b) are necessary conditions. It follows from Corollary 7.3 that H is the field, generated by $G^*(H)$, and L is generated by $G^*(L)$. Since $G^*(H)$ is the join of the groups $G^*_n(H)$ for n in e and $G^*(L)$ is the join of the groups $G^*_n(L)$ for n in e, it follows from (a) and (b) that H as well as L is a field, generated from K by solving all the equations

$$x^n - k = 0 \text{ for } n \text{ in } e \text{ and } k \text{ in } P(n).$$

The existence of a K-isomorphism between H and L is now a consequence of a well known theorem in the theory of fields.[10]

COROLLARY 8.3. *Suppose that H and L are algebraic, normal and separable extensions of the field K, that $\mathbf{R}(K < H)$ and $\mathbf{R}(K < L)$ are abelian, that (7.R) is satisfied by H and L, and that the positive integer n is the l. c. m. of the integers in $e(K < H)$. There exists a K-isomorphism between H and L if, and only if,*

(a′) *n is the l. c. m. of the integers in $e(K < L)$;*

(b′) $P(n, K < H) = P(n, K < L)$.

This is a consequence of the proof of Theorem 8.2, since the conditions concerning n imply that $G^*(F) = G^*_n(F)$ for $F = H, L$.

Remark. If the fields H and L are contained in a field M, e.g. in the algebraic closure of K, then the conditions of Theorem 8.2 or Corollary 8.3 imply not only the existence of a K-isomorphism between H and L, but even the *equality* of H and L, since by Lemma 6.3 every K-isomorphism of fields between K and its algebraic closure is induced by an automorphism of this algebraic closure, and since furthermore normal, algebraic extensions are mapped upon themselves by K-automorphisms of the algebraic closure of K.

THEOREM 8.4. *Suppose that e is a set of positive integers, containing all the divisors of its elements and the l. c. m. of every finite subset, that K is a field, containing n different n-th roots of unity for every n in e, and that $D(n)$ is—for every n in e—a multiplicative subgroup of K^*.*

There exists an algebraic, normal and separable extension F of K such that $\mathbf{R}(K < F)$ is abelian, $e(K < F) = e$ and $P(n, K < F) = D(n)$ for every n in e if, and only if,

(1) $K^{*n} \leq D(n)$;

[10] E. Steinitz, *loc. cit.*, p. 112.

(2) *$D(n)/K^{*n}$ contains elements of order n;*

(3) $D(n)^k \leq D(nk)$;

(4) *the elements of $D(nk)$ whose order mod K^{*nk} divides n are represented exactly by the elements of $D(n)^k$.*

Proof. A. Suppose that F is an algebraic, normal and separable extension of the field K, that (7. R) is satisfied and that $\boldsymbol{R}(K < F)$ is abelian. Then $K^{*n} \leq P(n, K < F)$, since $K^* \leq G^*{}_n(F)$. $e(K < F)$ contains exactly the orders of the elements in the character group of the vector group $\boldsymbol{R}(K < F)$. $P(n, K < F)/K^{*n}$ contains therefore elements of order n, if n is contained in $e(K < F)$, since $P(n, K < F)/K^{*n}$ and $G^*{}_n(F)/K^*$ are isomorphic and $G^*(F)/K^*$ is the character group of $\boldsymbol{R}(K < F)$. If u is an element of $P(n, K < F)$, then there exists an element v in F such that $v^n = u$. $u^k = v^{nk}$ is therefore an element of $P(nk, K < F)$. If u is an element of $P(nk, K < F)$ whose order mod K^{*nk} is a divisor of n, then there exists an element f in F such that $f^{nk} = u$ and an element h in K^* such that $u^n = h^{nk}$. Consequently $(f^n)^{nk} = u^n = h^{nk}$ and $f^n h^{-1}$ is an nk-th root of unity. Since the nk-th roots of unity are by (7. R) elements of K, it follows that f^n is an element of K and u is therefore an element of $P(n, K < F)^k$. If conversely the element u in $P(nk, K < F)$ is at the same time an element of $P(n, K < F)^k$, then $u = v^k$ for a suitable element v in k^* and $u^n = v^{nk}$ is therefore an element of K^{*nk}. Thus it has been proved that the conditions (1) to (4) are necessary conditions.

B. Assume now that the groups $D(n)$ satisfy the conditions (1) to (4). Then denote by F the essentially uniquely determined field which is generated from K by completely solving the equations

$$x^n - d = 0 \text{ for } n \text{ in } e \text{ and } d \text{ in } D(n),$$

and denote by G^* the set of all the solutions of these equations in F.

Since K contains for every n in e exactly n different n-th roots of unity, these numbers n are relatively prime to the characteristic of K and F is therefore[11] an algebraic, normal and separable extension of K.

If $u(i)$ is for $i = 1, 2$ an element of G^*, then there exists a number $n(i)$ in e such that $u(i)^{n(i)}$ is an element of $D(n(i))$ and there exists the l. c. m. m of $n(1)$ and $n(2)$ in e. Since $m = n(i)k(i)$, it follows that $u(i)^m = (u(i)^{n(i)})^{k(i)}$ is an element of $D(n(i))^{k(i)}$ and therefore by (3) an element of $D(m)$. Since $D(m)$ is a group, $(u(1)u(2))^m$ is an element of

[11] E. Steinitz, *loc. cit.*, p. 66.

$D(m)$ and $u(1)u(2)$ is consequently an element of G^*. G^* is therefore by condition (1) and by the definition of G^* a group between K^* and $G^*(F)$. Since F is generated by G^*, it follows from Theorem 7.1, (b) that

$$\boldsymbol{R}(K < F) \text{ is abelian, } F \text{ satisfies (7.R) and } G^* = G^*(F).$$

Since e is the set of orders of elements in G^*/K^*, it follows from Theorem 7.1 that $e = e(K < F)$.

It is an obvious consequence of the definition of F and G^* that $D(n) \leq P(n, K < F)$ for every n in e. Assume now that u is any element in $P(n, K < F)$. Then there exists an element v in $G^*(F)$ such that $v^n = u$. Since $G^* = G^*(F)$, there exists an integer m in e such that v^m is an element of $D(m)$. Denote by q the l. c. m. of n and m. q is an element of e and $q = nn' = mm'$. Now it follows from condition (3) that $v^q = (v^m)^{m'}$ is an element of $D(q)$ and, since the order mod K^{*m} of $v^q = u^{n'}$ is a divisor of n, it follows from condition (4) that $u^{n'} = v^q = w^{n'}$ for some w in $D(n)$. uw^{-1} is therefore an n'-th root of unity and thus contained in K, since n' is a divisor of q. Since the n'-th roots of unity are for the same reason elements of K^{*n}, it follows from (1) that $u = (uw^{-1})w$ is an element of $D(n)$. Hence

$$P(n, K < F) = D(n) \text{ for every } n \text{ in } e,$$

and this completes the proof of the Theorem.

Corollary 8.5. *Suppose that the field K contains n different n-th roots of unity and that P is a multiplicative group of elements $\neq 0$ in K. Then there exists an algebraic, normal and separable extension F of K such that $\boldsymbol{R}(K < F)$ is abelian, n is the l. c. m. of the numbers in $e(K < F)$ and $P(n, K < F) = P$ if, and only if $K^{*n} \leq P$ and P/K^{*n} contains elements of order n.*

Remark. In this exposition of a theory of abelian extensions it has been assumed that the set $e(K < F)$ does not contain the characteristic of the field K. If the characteristic of the field K is a prime number p and $e(K < F)$ consists of powers of p only, then it is again possible to extend the methods developed by Witt for the finite case to the general case. But a more detailed discussion of this case may be omitted, since it may be handled in exactly the same way and the translation from the case "prime to characteristic" to the other case follows exactly the lines, indicated by Witt.

The University of North Carolina,
Chapel Hill, N. C.

A NOTE ON TOPOLOGICAL FIELDS.*

By N. Jacobson.[1]

In a recent paper[2] the author showed that a locally compact, separable, totally disconnected (l. c. s. t. d.) field $\mathfrak{F}$ has a valuation determined as follows: Let $\mathfrak{J}$ be the totality of elements a, called integers, such that $\{a^\nu\}$ is bounded (contained in a compact and closed set)[3] and $\mathfrak{P}$ the subset of elements b such that $b^\nu \to 0$. Then $\mathfrak{J}$ is a compact and open domain of integrity and $\mathfrak{P}$ a two-sided principal prime ideal $= (x)$ in $\mathfrak{J}$. Every element y in $\mathfrak{F}$ has the form ux^k where k is a rational integer and $u \in \mathfrak{J} \mid \mathfrak{P}$ the part of $\mathfrak{J}$ not in $\mathfrak{P}$. If we define $\exp y$ to be k and $|y| = \gamma^{\exp y}$ where $0 < \gamma < 1$ then y is a non-archimedean valuation of $\mathfrak{F}$, i. e.,

$$|y_1 y_2| = |y_1|\ |y_2| \qquad |y_1 + y_2| \leqq \max(|y_1|, |y_2|).$$

We showed also that $\mathfrak{F}$ is a cyclic algebra over its centrum $\mathfrak{C}$ and the latter is either a $\mathfrak{p}$-adic field[4] or a field of power series in an indeterminate z with coefficients in a finite field. The former case obtains if the characteristic $\chi(\mathfrak{F}) = 0$ and the latter, which we shall call a z-adic field, if $\chi(\mathfrak{F}) = p \neq 0$.

Now if $\mathfrak{C}$ is any commutative $\mathfrak{p}$-adic or z-adic field it may be topologized by means of the neighborhoods $\{a + U_k\}$ of a where U_k is the set of points $(\alpha_0 + \alpha_1\pi + \cdots)\pi^k$, $k = 0, \pm 1, \pm 2, \cdots$, $(\mathfrak{p} = (\pi))$ in the $\mathfrak{p}$-adic field or $(\alpha_0 + \alpha_1 z + \cdots)z^k$ in the z-adic field. In this topology a sequence of elements $a^{(\nu)} = (\alpha_0^{(\nu)} + \alpha_1\pi^{(\nu)} + \cdots)\pi^{k_\nu}$ $((\alpha_0^{(\nu)} + \alpha_1^{(\nu)}z + \cdots)z^{k_\nu})$ converges to 0 if and only if $k_\nu \to \infty$. It follows that the neighborhoods $a + U_k$ are closed and, since the α_i have only a finite range, it is easily seen that they are also compact. The set of elements of finite length, e. g. $(\alpha_0 + \alpha_1\pi + \cdots \alpha_m\pi^m)\pi^k$ is denumerable and everywhere dense in $\mathfrak{C}$. Hence $\mathfrak{C}$ is a l. c. s. t. d. field. If $\mathfrak{F}$ is an algebra with a finite basis over $\mathfrak{C}$ it is a vector space over $\mathfrak{C}$ and hence it is l. c. s. t. d. in the usual vector space topology. Thus we see that the theory of l. c. s. t. d. fields is equivalent to that of division algebras over a $\mathfrak{p}$-adic or a z-adic field. In this note we shall apply the topological methods

* Presented to the Society September 7, 1937. Received by the Editors June 15, 1937.

[1] National Research Fellow.

[2] "Totally disconnected locally compact rings," *American Journal of Mathematics*, vol. 58 (1936), pp. 433-449. This paper will be referred to as T.

[3] An equivalent condition is that $\{a^\nu\}$ have no divergent subsequence: T, p. 444.

[4] i. e. an algebraic extension of finite order of a field of Hensel p-adic numbers.

of our earlier paper to investigate the automorphisms and anti-automorphisms of $\mathfrak{F}$.

Let J be an automorphism or an anti-automorphism of the l. c. s. t. d. field $\mathfrak{F}$, i. e. a $(1-1)$ correspondence of $\mathfrak{F}$ on itself such that $(a+b)^J = a^J + b^J$ and $(ab)^J = a^J b^J$ (automorphism) or $(ab)^J = b^J a^J$ (anti-automorphism). If $c \in$ the centrum $\mathfrak{C}$, $c^J \in \mathfrak{C}$ also and hence the correspondence $c \to c^J$ is an automorphism of $\mathfrak{C}$. By a theorem of F. K. Schmidt [5] J is a continuous transformation in $\mathfrak{C}$.

Suppose $b_1, b_2, \cdots, b_m$ $(m = r^2)$ [6] is a basis for $\mathfrak{F}$ over $\mathfrak{C}$. Then every y may be written uniquely in the form $\Sigma c_i b_i$, $c_i \in \mathfrak{C}$. Set $c_i = E_i(y)$.

LEMMA 1. *$E_i(y)$ is a continuous function of y.*

Since $E_i(y)$ is linear it suffices to show that if $y_\nu = \Sigma c_i^{(\nu)} b_i \to 0$ then $E_i(y_\nu) = c_i^{(\nu)} \to 0$. If $c_i^{(\nu)} \not\to 0$ there exists a subsequence which does not have 0 as a limit point and we may suppose that this is the whole sequence $\{c_i^{(\nu)}\}$. For each ν let $\lambda(\nu) = 1, \cdots, m$ be an index such that $| c_{\lambda(\nu)}^{(\nu)} | \geqq | c_j^{(\nu)} |$ for all $j = 1, \cdots, m$. Since $\lambda(\nu)$ has a finite range, one of its values, say $\lambda(\nu) = 1$, occurs infinitely often and hence by restricting ourselves to a subsequence we may suppose that $| c_1^{(\nu)} | \geqq | c_j^{(\nu)} |$ for all ν. Since $\lim c_i^{(\nu)} \neq 0$, $\lim c_1^{(\nu)} \neq 0$ and there is a subsequence such that $\lim (c_1^{(\nu_k)})^{-1} = d \in \mathfrak{C}$.[7] Again we write ν for ν_k Then

$$\lim (c_1^{(\nu)})^{-1} y_\nu = 0 = \lim (b_1 + (c_1^{(\nu)})^{-1} c_2^{(\nu)} b_2 + \cdots + (c_1^{(\nu)})^{-1} c_m^{(\nu)} b_m).$$

Since $|(c_1^{(\nu)})^{-1} c_j^{(\nu)} | \leqq 1$ we may suppose that $c_1^{(\nu)} c_j^{(\nu)} \to c_j \in \mathfrak{C}$. It follows that $b_1 + c_2 b_2 + \cdots + c_m b_m = 0$ contrary to the linear independence of the b's.

LEMMA 2. *J is a continuous mapping of $\mathfrak{F}$ on itself.*

Suppose

$$y_\nu = c_1^{(\nu)} b_1 + \cdots + c_m^{(\nu)} b_m \to y = c_1 b_1 + \cdots + c_m b_m.$$

By Lemma 1, $c_i^{(\nu)} \to c_i$ and since J is continuous in $\mathfrak{C}$, $c_i^{(\nu) J} \to c_i^J$. Hence

$$y_\nu^J = c_1^{(\nu) J} b_1^J + \cdots + c_m^{(\nu) J} b_m^J \to y^J = c_1^J b_1^J + \cdots + c_m^J b_m^J.$$

---

[5] F. K. Schmidt, "Mehrfach perfekte Körper," *Mathematische Annalen*, vol. 108 (1933), pp. 1-25.

[6] The order of a division algebra over its centrum is the square of an integer called its degree. See, for example, Deuring's "Algebren," *Ergebnisse der Mathematik*, p. 47.

[7] We require here the property of a locally compact field that $a_\nu \to 0$ if $\{a_\nu\}$ diverges: T, p. 442.

Since J^{-1} is also an automorphism or anti-automorphism it is continuous and hence J is a homeomorphism of $\mathfrak{F}$.

THEOREM 1. *Any automorphism or anti-automorphism J of $\mathfrak{F}$ is isometric, i. e., $|y^J| = |y|$ for every y in $\mathfrak{F}$.*

The condition that $\exp u = 0$ is that $\{u^\nu\}$ and $\{u^{-\nu}\}$ be bounded (or have no divergent subsequence). Since J is a homeomorphism $\{(u^J)^\nu\} = \{(u^\nu)^J\}$ and $\{(u^J)^{-\nu}\}$ are bounded also. Hence $\exp u^J = 0$. Now let x be an element such that $\mathfrak{P} = (x)$, or $\exp x = 1$. Consider x^J. Since $(x^J)^\nu = (x^\nu)^J \to 0, x^J \in \mathfrak{P}$ and hence $\exp x^J = l > 0$. Similarly $\exp x^{J^{-1}} = k > 0$. If $x^J = ux^l$ where $\exp u = 0$, $x = u^{J^{-1}}(x^{J^{-1}})^l$ or $(x^{J^{-1}})^l u^{J^{-1}}$. Hence $1 = \exp x = lk$ and so $l = k = 1$. Thus $\exp x^J = 1$. If $\exp y = n$ an integer, $y = ux^n$ and $y^J = u^J(x^J)^n$ or $= (x^J)^n u^J$ and hence $\exp y^J = n$. Hence for any y, $|y^J| = |y|$.

The condition that $\mathfrak{F}$ have an automorphism is evidently no restriction since the identity transformation is such a correspondence. On the other hand we shall see that the existence of an anti-automorphism strongly restricts the structure of $\mathfrak{F}$.

We recall that $\mathfrak{F}$ is cyclic over $\mathfrak{C}$. Its degree $r(= \sqrt{m})$ over $\mathfrak{C}$ may be characterized as the order of the automorphism $u \to xux^{-1} \pmod{\mathfrak{P}}$ of the finite field $\mathfrak{F} - \mathfrak{P}$ where x is determined as above.[8] We have for any u in $\mathfrak{F}$

$$xux^{-1} \equiv u^s \pmod{\mathfrak{P}} \qquad (s = p^t)$$

where p is the characteristic of $\mathfrak{F} - \mathfrak{P}$. It follows that

$$\begin{aligned}(x^J)^{-1}u^Jx^J &\equiv (u^s)^J \pmod{\mathfrak{P}}\\ &\equiv (u^J)^s \pmod{\mathfrak{P}}.\end{aligned}$$

Since $|x^J| = |x|$, $x^J = vx$ where $|v| = 1$. Then

$$(u^J)^s \equiv x^{-1}(v^{-1}u^Jv)x \equiv x^{-1}u^Jx \pmod{\mathfrak{P}}$$

since $v^{-1}u^Jv \equiv u^J \pmod{\mathfrak{P}}$. Thus

$$x^{-1}u^Jx \equiv xu^Jx^{-1} \pmod{\mathfrak{P}}$$

for all u^J. Since u^J varies over a complete set of residues mod $\mathfrak{P}$ when u does we have

$$\begin{aligned}x^{-1}ux &\equiv xux^{-1} \pmod{\mathfrak{P}}\\ x^2ux^{-2} &\equiv u \pmod{\mathfrak{P}}.\end{aligned}$$

Thus the order of the automorphism $u \to xux^{-1} \pmod{\mathfrak{P}}$ is either 1 or 2 and hence $r = 1$ or 2.

[8] T, pp. 446-448.

THEOREM 2. *The degree of an anti-automorphic l. c. s. t. d. field is either* 1 *or* 2.

The condition $r = 1$ or 2 is also sufficient that $\mathfrak{F}$ have an anti-automorphism. For if $r = 1$, $\mathfrak{F}$ is commutative and any automorphism is an anti-automorphism. If $r = 2$, $\mathfrak{F}$ is a quaternion algebra over its centrum and it is well known that $\mathfrak{F}$ has anti-automorphisms J. In fact J may be chosen so that $J^2 = I$ the identity and $c^J = c$ for all c in $\mathfrak{C}$.

We suppose now that J is an involutorial anti-automorphism $(J^2 = I)$ and the characteristic $\chi(\mathfrak{F}) \neq 2$. Let $\mathfrak{K}_J$ denote the symmetric or invariant elements $(h^J = h)$, $\mathfrak{S}_J$ the set of skew elements $(s^J = -s)$ and $\mathfrak{C}_0 = \mathfrak{K}_J \frown \mathfrak{C}$. Evidently $\mathfrak{C}_0$ is a closed subfield of $\mathfrak{C}$ and hence is l. c. s. t. d. For any y we have $y = \frac{1}{2}(y + y^J) + \frac{1}{2}(y - y^J) = h + s$ where $h = \frac{1}{2}(y + y^J) \in \mathfrak{K}_J$ and $s = \frac{1}{2}(y - y^J) \in \mathfrak{S}_J$. In particular this holds for y in $\mathfrak{C}$ and in this case $h \in \mathfrak{C}_0$, $s \in \mathfrak{S}_J \frown \mathfrak{C}$. Hence if $\mathfrak{C} \neq \mathfrak{C}_0$ there exists an element $q \neq 0$ in $\mathfrak{S}_J \frown \mathfrak{C}$. Then $y = h + (sq^{-1})q$ and $sq^{-1} \in \mathfrak{K}_J$. For c in $\mathfrak{C}$ we obtain a quadratic extension $\mathfrak{C}_0(q)$ of $\mathfrak{C}_0$. It follows that $\mathfrak{C}_0$ is infinite and therefore is either a $\mathfrak{p}$-adic or a z-adic field.[9] We note also that $q^2 = q_0$ is symmetric

We wish to show that if $\mathfrak{C} \neq \mathfrak{C}_0$ then $\mathfrak{F} = \mathfrak{C}$, i. e., $\mathfrak{F}$ is commutative. For the present we drop the restriction that $\mathfrak{C}_0$ is l. c. s .t. d. and allow it to denote any commutative field of characteristic $\neq 2$. We suppose that $\mathfrak{F}$ is a division algebra of order 8 over $\mathfrak{C}_0$ with centrum $\mathfrak{C} = \mathfrak{C}_0(q)$ and J is an involutorial anti-automorphism of $\mathfrak{F}$ leaving the elements of $\mathfrak{C}_0$ invariant and mapping q into $-q$. The order of $\mathfrak{F}$ over $\mathfrak{C}$ is of course 4.

It is easily seen that $\mathfrak{K}_J$ and $\mathfrak{S}_J$ defined as above are vector spaces over $\mathfrak{C}_0$. If $h \in \mathfrak{K}_J$, $hq \in \mathfrak{S}_J$ and if $s \in \mathfrak{S}_J$, $sq \in \mathfrak{K}_J$. Hence these spaces have the same order over $\mathfrak{C}_0$. Evidently $\mathfrak{S}_J \frown \mathfrak{K}_J = 0$ and as we showed before $\mathfrak{F} = \mathfrak{S}_J + \mathfrak{K}_J$. It follows that $\mathfrak{K}_J$ and $\mathfrak{S}_J$ have order 4 over $\mathfrak{C}_0$ and since $\mathfrak{K}_J = \mathfrak{S}_J q$, $\mathfrak{F} = \mathfrak{S}_J \mathfrak{C}$.[10]

Let $T(y)$ and $N(y)$ denote respectively the trace and norm of y relative to $\mathfrak{C}$. Then $y^2 - T(y)y + N(y) = 0$ and if $y \notin \mathfrak{C}$ this is the equation of least degree having coefficients in $\mathfrak{C}$ and satisfied by y. Then

$$(y^J)^2 - T(y)^J y^J + N(y)^J = 0 = (y^J)^2 - T(y^J)y^J + N(y^J)$$

and hence for $y \notin \mathfrak{C}$

[9] We note that any infinite closed subfield $\mathfrak{C}_0$ of a l. c. s. t. d. field is not discrete. For if $\{a_\nu\}$ is a divergent sequence in $\mathfrak{C}_0$ then $a_\nu^{-1} \to 0$.

and hence $\in \mathfrak{C}_0$.

[10] If $\mathfrak{A}$ and $\mathfrak{B}$ are vector spaces over $\mathfrak{C}_0$, $\mathfrak{A}\mathfrak{B}$ denotes the smallest vector space over $\mathfrak{C}_0$ containing all the products ab where $a \in \mathfrak{A}$ and $b \in \mathfrak{B}$.

$$(*) \qquad T(y^J) = T(y)^J \qquad N(y^J) = N(y)^J.$$

But for $y \in \mathfrak{C}$, $T(y) = 2y$, $N(y) = y^2$ and so (*) holds for every y in $\mathfrak{F}$. If $y \in \mathfrak{S}_J$, $T(y)^J = T(y^J) = T(-y) = -T(y)$ and $N(y)^J = N(-y) = N(y)$, i.e., $T(y) \in \mathfrak{S}_J \frown \mathfrak{C}$ and $N(y) \in \mathfrak{C}_0$. $\mathfrak{S}_J \frown \mathfrak{C}_0$ consists of the $\mathfrak{C}_0$-multiples of q.

Let $\mathfrak{S}'_J$ be the totality of elements of trace 0 in $\mathfrak{S}_J$. $\mathfrak{S}'_J$ is a vector space over $\mathfrak{C}_0$ and if $y, t \in \mathfrak{S}'_J$ so does $[y, t] = yt - ty$. If y is any element of $\mathfrak{S}_J$, $y = \frac{1}{2}T(y) + y_0$ where $y_0 = y - \frac{1}{2}T(y) \in \mathfrak{S}'_J$. Thus $\mathfrak{S}_J = \mathfrak{S}'_J + \mathfrak{C}_0 q$ and $\mathfrak{S}'_J \frown \mathfrak{C}_0 q = 0$. It follows that the order of $\mathfrak{S}'_J$ over $\mathfrak{C}_0$ is 3. Since $1 \not\in \mathfrak{C}'_J$, the order of $\mathfrak{F}_0 = \mathfrak{S}'_J + \mathfrak{C}_0 1$ is 4 over $\mathfrak{C}_0$.

For y, t in $\mathfrak{S}'_J$ we have

$$\begin{aligned} yt &= \tfrac{1}{2}[(y+t)^2 - y^2 - t^2 + yt - ty] \\ &= \tfrac{1}{2}[-N(y+t) + N(y) + N(t) + yt - ty] \in \mathfrak{F}_0. \end{aligned}$$

Hence $\mathfrak{F}_0$ is an algebra over $\mathfrak{C}_0$. Since $\mathfrak{F}_0 = \mathfrak{S}'_J + \mathfrak{C}_0 1$, $\mathfrak{S}_J = \mathfrak{S}'_J + \mathfrak{C}_0 q$ and $\mathfrak{F} = \mathfrak{S}_J \mathfrak{C}$, we have $\mathfrak{F} = \mathfrak{F}_0 \mathfrak{C}$. We have therefore proved the following theorem.

THEOREM 3. *If $\mathfrak{F}$ is a division algebra of order 8 over $\mathfrak{C}_0$ of characteristic $\neq 2$ with $\mathfrak{C} = \mathfrak{C}_0(q)$, $q^2 \in \mathfrak{C}_0$, as centrum and J an involutorial anti-automorphism in $\mathfrak{F}$ such that $q^J = -q$ then $\mathfrak{F}$ is a direct product $\mathfrak{C} \times \mathfrak{F}_0$ of $\mathfrak{C}$ and $\mathfrak{F}_0$ the join of $\mathfrak{C}_0 1$ and the space of skew elements of trace 0.*[11]

Since the centrum of $\mathfrak{F}$ is $\mathfrak{C}$ that of $\mathfrak{F}_0$ must be $\mathfrak{C}_0$.

If $\mathfrak{C}_0$ is a commutative l. c. s. t. d. field it is well known that an algebra $\mathfrak{F}_0$ of degree 2 over $\mathfrak{C}_0$ has every quadratic field as a splitting field, i. e., if $\mathfrak{C} = \mathfrak{C}_0(q)$ then $\mathfrak{C} \times \mathfrak{F}_0$ is a complete matrix algebra over $\mathfrak{C}$.[12] Hence

THEOREM 4. *If $\mathfrak{F}$ is an l. c. s. t. d. field of characteristic $\neq 2$ and J an involutorial anti-automorphism such that $c^J \neq c$ for some c in the centrum $\mathfrak{C}$, then $\mathfrak{F} = \mathfrak{C}$ is commutative.*

In the remainder of the paper we shall extend the above results to $\mathfrak{F}_n$ a complete matrix algebra of n rows and columns over an l. c. s. t. d. field $\mathfrak{F}$. Let e_{ij} $(i, j = 1, \cdots, n)$ be a set of matrix units in $\mathfrak{F}_n$; $e_{ij}e_{kl} = \delta_{jk}e_{il}$ and every element b in $\mathfrak{F}_n$ has the form $\Sigma y_{ij}e_{ij}$ where $y_{ij} \in \mathfrak{F}$. As before let $\mathfrak{C}$

[11] Cf. A. A. Albert, "Involutorial simple algebras and real Riemann matrices," *Annals of Mathematics*, vol. 36 (1935), p. 909.

[12] Deuring's *Algebren*, p. 113. The proof given there for $\mathfrak{C}_0$ a $\mathfrak{p}$-adic field holds also for a z-adic field. See *Algebren*, p. 137.

denote the centrum of $\mathfrak{F}$ or of $\mathfrak{F}_n$. We recall that $\mathfrak{F}$ may be characterized as the totality of elements commutative with the e_{ij}.

Let J be an anti-automorphism of $\mathfrak{F}_n$ and $e_{ji}{}^J = f_{ij}$. Then $f_{ij}f_{kl} = \delta_{jk}f_{il}$ and it follows that there exists a non-singular element s in $\mathfrak{F}_n$ such that $e_{ij} = s^{-1}f_{ij}s$. The correspondence $b \to b^K = s^{-1}b^J s$ is an anti-automorphism also and $e_{ij}{}^K = e_{ji}$. Hence if $y \in \mathfrak{F}$, y^K which commutes with the e_{ij} belongs to $\mathfrak{F}$ also and $\mathfrak{F}$ itself is anti-automorphic. By Theorem 1 we have

THEOREM 5. *If $\mathfrak{F}$ is an l. c. s. t. d. field such that $\mathfrak{F}_n$ is anti-automorphic then $\mathfrak{F}$ has degree* 1 *or* 2 *over its centrum.*

If J is involutorial, then

$$e_{ij} = e_{ij}{}^{J^2} = f_{ji}{}^J = (se_{ji}s^{-1})^J = (s^{-1})^J s e_{ij} s^{-1} s^J.$$

Thus $s^{-1}s^J$ commutes with all e_{ij} and $s^J = su$ where $u \in \mathfrak{F}$. If $u = -1$ it is easily seen that the correspondence $b \to b^K = s^{-1}b^J s$ is an involutorial anti-automorphism in $\mathfrak{F}_n$ mapping $\mathfrak{F}$ into itself. If $u \neq 1$, $s^J + s = s(u+1) = t$ is non-singular and $t^J = t$. Hence the correspondence $b \to b^K = t^{-1}b^J t$ is an involutorial anti-automorphism mapping $\mathfrak{F}$ into itself.[13] In either case $\mathfrak{F}$ is involutorial anti-automorphic. Evidently if $c^J \neq c$ then $c^K \neq c$ for c in $\mathfrak{C}$. Hence by Theorem 4 we have

THEOREM 6. *If $\mathfrak{F}$ is an l. c. s. t. d. field of characteristic $\neq 2$ and $\mathfrak{F}_n$ has an involutorial anti-automophism J mapping an element c of the centrum into $c^J \neq c$ then $\mathfrak{F} = \mathfrak{C}$ is commutative.*

PRINCETON UNIVERSITY AND
THE INSTITUTE FOR ADVANCED STUDY.

---

[13] Cf. Albert, *loc. cit.* in 11, p. 897.

THE QUATERNIONIC CONGRUENCE $\bar{t}at \equiv b \pmod{g}$ AND THE EQUATION $h(8n+1) = x_1^2 + x_2^2 + x_3^2$.*

By Gordon Pall.

1. A quaternion will be called integral if its coördinates are rational integers. Such quaternions will be represented by the letters $a, \cdots, e$, $t, \cdots, z$, and their integer coördinates will be distinguished by subscripts; thus $t = t_0 + it_1 + jt_2 + kt_3$.

If g is a rational integer, the congruence $c \equiv d \pmod{g}$ is equivalent to the system $c_f \equiv d_f \pmod{g}$ $(f = 0, 1, 2, 3)$. Most of the familiar rules for handling congruences can be extended, with due regard to lack of commutativity, to such congruences. For example, if $a \equiv b$ and $c \equiv d \pmod{g}$, then $ac \equiv bd$; for $ac - bd = (a-b)c + b(c-d)$. Again, if $Nx (= \bar{x}x = \Sigma x_f^2)$ is prime to g, $ax \equiv bx$ implies $a \equiv b$; for $(ax - bx)\bar{x} = (a-b)Nx$.

Some of our results (e. g. § 2 and (19)) will bear incidentally on cancellation of an inner factor. For instance, by (12) either of $\bar{t}it \equiv \pm i \pmod{4}$ implies $\bar{t}t \equiv 1 \pmod{4}$.

The study of the congruence $\bar{t}at \equiv b \pmod{g}$ is reducible to the case where a and b are pure quaternions (§ 5). If $g = 2^n$ and $4 \nmid Nb$, the conditions for mere solvability are simple (Theorems 2 and 4). But all solutions t satisfy one and only one of $\bar{t}t \equiv 1$ or $3 \pmod{4}$, as set forth in Theorems 3 and 5.

Our applications to the equation

$$h(8n+1) = x_1^2 + x_2^2 + x_3^2 \tag{1}$$

employ properties of the residues $\bar{t}at$ for moduli g dividing $4Na$, where a is any given pure quaternion $ia_1 + ja_2 + ka_3$. Analogous results, in part, exist for any g but will not be treated here.

For any g and (pure) a we obtain a system of residues $\bar{t}at \pmod{g}$ as t ranges over all integral quaternions. We regard the quaternions formed from a by permuting or changing the signs of a_1, a_2, a_3 as constituting a *set* $\Sigma(a)$. If a' is so formed by an even number of permutations and an even number of sign-changes, then for any odd modulus g, the residue-systems $\bar{t}at$ and $\bar{u}a'u$ are identical (§ 11, end). Thus there are associated with $\Sigma(a)$ at most four residue-systems $\pmod{g}$, g being odd, namely those represented by

* Received June 18, 1937.

$$(1') \qquad \bar{t}at,\ \bar{t}(-a)t,\ \bar{t}a^*t,\ \bar{t}(-a^*)t,$$

where $a^* = ia_1 + ja_3 + ka_2$. The four need not be distinct (Th. 9).

If $\bar{t}at \equiv b \pmod{g}$, $Nb \equiv Na(Nt)^2 \pmod{g}$. We now assume $g \mid Na$, g odd, a proper and pure. If g is a prime p, then $\bar{t}at$ represents besides $0 \pmod{p}$ exactly half the remaining $p^2 - 1$ residues b such that $p \mid Nb$; if $p \not\equiv 1 \pmod{8}$ the other half are represented by $\bar{t}(-a)t$, $\bar{t}a^*t$, or $\bar{t}(-a^*)t$ (§ 14). These results are extended to composite g (§ 14). If g contains a non-square factor $\equiv 1 \pmod{8}$ not all residues b of norm divisible by g are represented by (1′). It may be possible to supply the deficiency by other quaternions of the same norm as a, but this we shall not here investigate.

In (1) h and n are assumed to be non-negative integers, $4 \nmid h$. If $8n + 1$ is a square m^2, $m > 0$, then all proper integral solutions $x = ix_1 + jx_2 + kx_3$ of (1) are of the form $\bar{t}at$, where $a = ia_1 + ja_2 + ka_3$ is an integral solution of

$$(2) \qquad h = a_1{}^2 + a_2{}^2 + a_3{}^2$$

and t is a proper quaternion of norm m (*proper* meaning having coprime coördinates). To prove this we use the following well-known and fundamental result in the arithmetic of quaternions:

A proper quaternion, with norm divisible by an odd positive integer m, has exactly eight right-divisors (left-divisors) t of norm m. These form a class of left-associate (right-associate) quaternions $\pm t$, $\pm it$, $\pm jt$, $\pm kt$ ($\pm \bar{t}$, $\pm \bar{t}i$, $\pm \bar{t}j$, $\pm \bar{t}k$).

Consider then a proper, pure quaternion x of norm hm^2. Its left-divisors of norm m (unique up to a right unit factor) are the conjugates of the right-divisors. For if $x = vt$, $\bar{x} = -x = \bar{t}\bar{v}$. Hence, since $Nv = hm$, $\bar{t}$ is a left-divisor of v, $v = \bar{t}a$, $x = \bar{t}at$.

If x is derived in this way from a, $\Sigma(x)$ is derived from $\Sigma(a)$. For any proper, pure x of norm hm^2, $x = \bar{t}at$, where $Nt = m$, and t is unique up to a left unit factor. If t is replaced by $\pm it$, a is replaced by $-iai = ia_1 - ja_2 - ka_3$; similarly for $\pm jt$ and $\pm kt$. Evidently $-x$ is derived from $-a$; and if x_1, x_2, x_3 be permuted or changed in sign, so that $x \to x'$, then x' is derived from a' similarly obtained (cf. § 9, near end). Thus $\Sigma(x)$ is derived from $\Sigma(a)$. In § 15 we shall complete the proof of

THEOREM 1. *Let $h > 0$, $4 \nmid h$. Let h' be a positive divisor of h such that $h/h' \equiv 1 \pmod{8}$ (e. g. $h' = h$). Let A denote the class of all residues $\pmod{4h'}$ of proper solutions x of (1) derived from a single proper set of solutions of (2), with $8n + 1 = m^2$ ($m = 1, 5, 9, 13, \cdots$); and let B denote*

the class of residues $(\bmod 4h')$ *obtained similarly with* $m = 3, 7, 11, 15, \cdots$. *Then:*

I. the classes A *and* B *are mutually exclusive; and*

II. if h' *has no factor* > 1 *of the form* $8f+1$, *then all proper solutions of* (1) *belong to* A *or* B, *and for any* n *such that* $(8n+1)h/h'$ *is not a square,* (1) *has equally many solutions in each class* A, B.

Whenever (2) has only one proper set of solutions, e. g. when $h < 17$, the clause "derived from . . . of (2)" can be omitted. The classes A and B are then easily calculated by forming solutions of $h(4f \pm 1)^2 = x_1^2 + x_2^2 + x_3^2$, the work being abbreviated by using § 16. We give a table, listing only the least absolute residues $(\bmod 4h)$ of (x_1, x_2, x_3), and arranging each triple in a definite order:

h	residues $A, m = 4f+1$	residues $B, m = 4f+3$
1	(1, 0, 0)	(1, 2, 2)
2	(1, 1, 0), (3, 3, 4)	(3, 3, 0), (1, 1, 4)
3	(1, 1, 1), (1, 5, 5)	(1, 1, 5), (5, 5, 5)
5	(1 or 9, 0, 2), (3 or 7, 4, 10), (5, 8, 6)	(9 or 1, 10, 8), (7 or 3, 6, 0), (5, 2, 4)
6	(2, 1, 1), (2, 7, 7), (2, 5, 11)	(10, 1, 1), (10, 7, 7), (10, 5, 11)
	(10, 5, 5), (10, 11, 11), (10, 1, 7)	(2, 5, 5), (2, 11, 11), (2, 1, 7)
10	(0, 1 or 9, 3 or 13), (20, 7 or 17, 11 or 19)	change the even
	(4, 3 or 13, 15), (16, 7 or 17, 5)	or odd x_i
	(8, 11 or 19, 15), (12, 1 or 9, 5)	of A
11	(1, 1, 3), (3, 3, 13), (5, 5, 15), (7, 7, 21), (9, 9, 17)	to
	(21, 21, 3), (19, 19, 13), (17, 17, 15), (15, 15, 21), (13, 13, 17)	$2h - x_i$
	(1, 21, 19), (3, 19, 9), (5, 17, 7), (7, 15, 1), (9, 13, 5)	according as
13	(0, 2, 3 or 23), (20, 2, 5 or 21), (4, 14, 3 or 23),	$h \not\equiv 3$ or
	(12, 26, 5 or 21), (8, 14, 13), and multiply	$h \equiv 3 \pmod 8$.
	these by 3 and 9, reducing mod 52.	

For example let $h = 2$. If $8n+1$ is a square $(4f+1)^2$, $f \geqq 0$, the proper solutions of (1) have the two odd $x_i \equiv \pm 1 \pmod 8$ if the even x_j is $\equiv 0 \pmod 8$, but the two odd x_i are $\equiv \pm 3 \pmod 8$ if the even x_j is $\equiv 4 \pmod 8$; vice versa if $8n+1 = (4f+3)^2$. But if $8n+1$ is not a square then $2(8n+1)$ has equally many proper representations of either type. Thus $34 = 5^2 + 3^2 + 0^2 = 3^2 + 3^2 + 4^2$.

The case $h = 2$ can be given a more elegant form:

In the equation $2(8n+1) = x_1^2 + x_2^2 + x_3^2$ (x_3 even), there are equally many solutions of each type

$$P: \quad x_1 \equiv \pm x_2 \pmod{16}, \qquad Q: \quad x_1 \equiv \pm x_2 + 8 \pmod{16},$$

when $8n+1$ is not a square. If $8n+1=s^2$, $s>0$, then all proper solutions are of type P if $s \equiv 1$ or $3 \pmod 8$, but of type Q if $s \equiv 5$ or $7 \pmod 8$.

For A yields P if n is even, Q if n is odd; vice versa for B.

Hence $8n+1=y_1^2+y_2^2+2y_3^2$, where

$$y_1=\tfrac{1}{2}(x_1+x_2), \; y_2=\tfrac{1}{2}(x_1-x_2), \; x_3=2y_3.$$

As a corollary of cases $h=1$ and 2, every $8n+1$ is represented in both forms

$$x_1^2+16x_2^2+16x_3^2, \quad x_1^2+8x_2^2+64x_3^2;$$

whence both forms are easily shown to be regular, although their genera contain two classes. The present theory has extensions to many ternary quadratic forms.

As a further corollary, every $8n+1$ except 1 and 25, and every $2(8n+1)$ except 2, is a sum of three positive squares.[1]

For $h=1$ and 3 Theorem 1 is essentially equivalent to certain results of Jacobi and Glaisher obtained by theta-function expansions. For references, see Dickson's *History*, II, to Jacobi, pp. 262-3; Catalan, p. 266; Glaisher, p. 268; Bachmann, p. 270. Glaisher conjectured (*Quarterly Journal of Mathematics*, 20 (1885), p. 96) that his results could be proved by "actual transformation of the squares," and they were in fact proved recently by the writer by using automorphs of $x_1^2+x_2^2+x_3^2$, but not published. The close connection between such automorphs and quaternions is well known.

2. Even if Na is prime to g, $\bar{t}at \equiv a \pmod g$ need not imply $\bar{t}t \equiv 1 \pmod g$. For example, if g is *odd* and $a=i$, $\bar{t}it \equiv i$ expands into

$$t_0^2+t_1^2-t_2^2-t_3^2 \equiv 1, \quad -t_0t_3+t_1t_2 \equiv 0, \quad t_0t_2+t_1t_3 \equiv 0.$$

These are compatible with $\bar{t}t \equiv -1$ if $0 \equiv t_0 \equiv t_1 \equiv 1+t_2^2+t_3^2$, and with $\bar{t}t \equiv 1$ if $0 \equiv t_2 \equiv t_3 \equiv t_0^2+t_1^2-1$, either set of conditions being solvable.

That $\bar{t}t \equiv \pm 1$ are the only possibilities is shown in § 3, provided g is the power of a prime. We assume that g is the power of a prime throughout § 3.

3. On taking norms. From $\bar{t}at \equiv b \pmod g$ follows $\bar{t}at=b+gv$, $\bar{t}\bar{a}t=\bar{b}+g\bar{v}$, $\bar{t}at\bar{t}\bar{a}t=b\bar{b}+g(b\bar{v}+v\bar{b})+g^2v\bar{v}$, i. e.

[1] This was conjectured by the writer in "On Sums of Squares," *American Mathematical Monthly*, XL (1933), 10-18, on p. 11. The number 137 should be omitted from Theorem 3 of that article.

$$(3) \qquad (Nt)^2Na - Nb = g(b\bar{v} + v\bar{b}) + g^2Nv.$$

Here $b\bar{v} + v\bar{b}$, a quaternion plus its conjugate, is an even integer.

Assuming $Na \equiv Nb \pmod{2g}$ we draw some conclusions from (3). If Na is prime to g and g is odd, $(Nt)^2 \equiv 1$ and $Nt \equiv \pm 1 \pmod{g}$. We have just observed that both residues ± 1 are obtainable for Nt when $a = i$ and g is odd. Although, as we shall see for example in (19), the like result does not hold for even g, (3) yields only $Nt \equiv \pm 1 \pmod{g}$ if Na is prime to g (g even); and $Nt \equiv \pm 1 \pmod{\frac{1}{2}g}$ if $4 \mid g$, $Na \equiv 2 \pmod{4}$, $\frac{1}{2}Na$ prime to g. Conversely, if neither Na nor Nb is divisible by 4, and $2 \mid g$, (3) requires Nt odd, $Na \equiv Nb \pmod{4}$ if $2 \mid g$, $Na \equiv Nb \pmod{8}$ if $4 \mid g$. And if $8 \mid g$, and $Na \equiv Nb \equiv \pm 2 \pmod{8}$, (3) implies $Na \equiv Nb \pmod{16}$.

4. On expanding $\bar{t}at$ (cf. (42) and § 5) we see that

$$(4) \qquad \bar{t}at \equiv (Nt)a \pmod{2},$$

for any integral quaternions t, a. Hence if $2 \nmid b$,

$$\bar{t}at \equiv b \pmod{2} \text{ implies } a \equiv b \pmod{2}.$$

5. Henceforth we shall treat $\bar{t}at \equiv b \pmod{g}$ with $a = ia_1 + ja_2 + ka_3$, $b = ib_1 + jb_2 + kb_3$ (pure quaternions). Nothing essential is hereby lost. For if $x = a_0 + a$, $y = b_0 + b$, then the real part of $\bar{t}xt$ is $\bar{t}a_0t = a_0Nt$, and $\bar{t}xt \equiv y$ breaks up into

$$a_0Nt \equiv b_0, \qquad \bar{t}at \equiv b.$$

6. Principal theorems for $g = 2^n$. Hypothesis throughout:

$$(5) \qquad a = ia_1 + ja_2 + ka_3,\ b = ib_1 + jb_2 + kb_3$$

are pure integral quaternions, neither divisible by 2;

$$(6) \qquad l = 1 \text{ if } Na \text{ is odd}, \qquad l = 2 \text{ if } Na \equiv 2 \pmod{4}.$$

Like a familiar theorem on quadratic congruences in elementary theory of numbers is the following:

THEOREM 2. *The solvability for t, necessarily of odd norm, of*

$$(7) \qquad \bar{t}at \equiv b \pmod{4l}$$

implies, for every integer $n > 0$, the solvability for t of

$$(8) \qquad \bar{t}at \equiv b \pmod{2^n}.$$

However a new feature is presented in

THEOREM 3. *If* (7) *is solvable, then all solutions* t *satisfy one and only one of the congruences*

(9) $\qquad A:\ \bar{t}t \equiv 1 \pmod 4, \qquad B:\ \bar{t}t \equiv -1 \pmod 4.$

THEOREM 4. *Congruence* (7) *is solvable for* t *if and only if*

(10) $\qquad a \equiv b \pmod 2, \qquad Na \equiv Nb \pmod{8l}.$

THEOREM 5. *Conditions* (10) *being assumed, the solutions of* (7) *are of type* A *or* B *in accordance with the following rules:*

(11) *if* $Na \equiv 3 \pmod 8$, A *holds or* B *according as* $a_h \equiv -b_h \pmod 4$ *for an even* (0 *or* 2) *or odd* (1 *or* 3) *number of values* $h = 1, 2, 3$;

(12) *if* $Na \equiv 1 \pmod 4$, *and* a_1 (*say*) *is odd, then* A *holds if* $a_2 \equiv b_2$, $a_3 \equiv b_3$, B *if* $a_2 \equiv b_2 + 2$, $a_3 \equiv b_3 + 2 \pmod 4$;

(13) *if* $Na \equiv 2 \pmod 8$ *and* $4 | a_3$, A *holds if* $(b_1, b_3) \equiv (\pm a_1, a_3)$ *or* $(\pm 3a_1, a_3 + 4) \pmod 8$, B *if* $(b_1, b_3) \equiv (\pm a_1, a_3 + 4)$ *or* $(\pm 3a_1, a_3)$;

(13′) *if* $Na \equiv 6 \pmod 8$ *and* $a_3 \equiv b_3 \equiv \pm 2$, A *holds if* $\pm(b_1, b_2) \equiv (a_1, a_2)$ *or* $3(a_1, -a_2)$, B *if* $\pm(b_1, b_2) \equiv (a_1, -a_2)$ *or* $3(a_1, a_2)$; A *and* B *being reversed if* $a_3 \equiv -b_3 \pmod 8$.

The subscripts 1, 2, 3 *may be permuted in* (12)-(13′).

Theorem 2 is proved in § 7, 3 in § 8, 4 and 5 in §§ 9, 10.

7. Proof of Theorem 2. Proceeding by induction we may suppose that $\bar{t}at - b = 2^n w$, where $w = iw_1 + jw_2 + kw_3$ is a pure, integral quaternion, $n > l$.

First take Na to be odd, $n \geqq 2$, and set

(14) $\qquad u = t + 2^{n-1}z$, z an integral quaternion.

Then $\bar{u}au - b = 2^n e$, where

$$e = w + \tfrac{1}{2}(\bar{z}at + \bar{t}az) + 2^{n-2}\bar{z}az,$$

and we wish to choose z to make e even. As in (4), $\bar{z}az \equiv (Nz)a \pmod 2$, so that it suffices to have $\frac{1}{2}(\bar{t}az + \bar{z}at) \equiv w \pmod 2$ and Nz even. Abbreviate $\bar{t}a = d = d_0 + id_1 + \cdots$, whence $at = -\bar{d}$ and Nd is odd. The desired conditions expand into

$$(15)\qquad \begin{aligned} d_1z_0 + d_0z_1 - d_3z_2 + d_2z_3 &\equiv w_1,\\ d_2z_0 + d_3z_1 + d_0z_2 - d_1z_3 &\equiv w_2,\\ d_3z_0 - d_2z_1 + d_1z_2 + d_0z_3 &\equiv w_3,\\ -z_0 + z_1 - z_2 + z_3 &\equiv 0, \pmod 2). \end{aligned}$$

The determinant of the coefficients in the left members is

$$Nd(d_0 + d_1 + d_2 + d_3).$$

This being odd, conditions (15) can be satisfied.

Second take $Na \equiv 2 \pmod 4$, $n \geqq 3$, and set

$$(16)\qquad u = t + 2^{n-2}\theta z,$$

where $\theta = 1 + i$, $1 + j$, or $1 + k$ according as a_1, a_2, or a_3 is even. Then $\bar{\theta}a = 2v$, where v is an integral quaternion of odd norm. Now $\bar{u}au - b = 2^n e$, where $e = w + \frac{1}{2}(\bar{z}vt - \bar{t}\bar{v}z) + 2^{n-3}\bar{z}v\theta z$. Since Nv is odd, the argument proceeds as in the first case.

8. We prove a more general result than Theorem 3:

THEOREM 6. *Assume* (5) *and* (6), $n \geqq 2$. *All solutions* t *of*

$$(17)\qquad \bar{t}at \equiv b \pmod{l \cdot 2^n},$$

if any exist, satisfy one and only one of

$$(18)\qquad \bar{t}t \equiv 1 \pmod{2^n}, \qquad \bar{t}t \equiv -1 \pmod{2^n},$$

the particular one depending only on the residues of a *and* b $\pmod{4l}$.

It is expedient first to prove:

$$(19)\qquad \bar{t}at \equiv a \pmod{l \cdot 2^n} \text{ implies } \bar{t}t \equiv 1 \pmod{2^n};$$

(5) and (6) being assumed. Since $\bar{t}at \equiv a$, $(Nt)at \equiv ta$, i. e.

$$(20)\qquad Nt(a_1t_1 + a_2t_2 + a_3t_3) \equiv a_1t_1 + a_2t_2 + a_3t_3,$$

$$(21)\qquad Nt(a_1t_0 + a_2t_3 - a_3t_2) \equiv a_1t_0 - a_2t_3 + a_3t_2,$$

$$(22)\qquad Nt(-a_1t_3 + a_2t_0 + a_3t_1) \equiv a_1t_3 + a_2t_0 - a_3t_1,$$

$$(23)\qquad Nt(a_1t_2 - a_2t_1 + a_3t_0) \equiv -a_1t_2 + a_2t_1 + a_3t_0,$$

to modulus $l \cdot 2^n$. Form $(21)a_1 + (22)a_2 + (23)a_3$:

$$(24)\qquad (Nt)t_0(a_1^2 + a_2^2 + a_3^2) \equiv t_0(a_1^2 + a_2^2 + a_3^2), \pmod{l \cdot 2^n}).$$

The conclusion $\bar{t}t \equiv 1 \pmod{2^n}$ follows from (20) or (24), if

(25) $$a_1t_1 + a_2t_2 + a_3t_3 \text{ or } t_0 \text{ is odd.}$$

In fact (25) always holds in consequence of (20)-(23): to see this, we prove the impossibility of

(α) t_0 even, t_1, t_2, t_3 odd, a_1, a_2 odd, a_3 even;
(β) t_0, t_1, t_2 even, t_3 odd, a_3 even.

In case (α), $Na \equiv 2 \pmod 4$, and (21) holds to modulus 8. If $4|t_0$, $Nt \equiv 3 \pmod 8$, (21) reduces to $4 \equiv 0 \pmod 8$. If $4|t_0 - 2$, $Nt \equiv -1 \pmod 8$, (21) implies $8|2a_1t_0$. In case (β), $Nt \equiv 1 \pmod 4$, and (21) yields $4|2(a_2t_3 - a_3t_2)$, a contradiction unless a_2 is even. But then a_1 is odd, and (22) yields $4|2(a_3t_1 - a_1t_3)$, whereas $a_3t_1 - a_1t_3$ is odd.

Now Theorem 6 will follow from Theorem 3. For by § 3, (17) implies $\bar{t}t \equiv \pm 1 \pmod{2^n}$, and only one of these agrees with either residue $\pm 1 \pmod 4$.

If, besides (7) we have $\bar{u}au \equiv b \pmod{4l}$, then $(Nu)^2a \equiv ub\bar{u}$, $u\bar{t}at\bar{u} \equiv a \pmod{4l}$, and by (19) with $n = 2$, $N(t\bar{u}) = NtNu \equiv 1 \pmod 4$, $Nt \equiv Nu \pmod 4$.

9. Let us write aAb to indicate that (7) is solvable with $\bar{t}t \equiv 1$, and aBb to express solvability with $\bar{t}t \equiv -1 \pmod 4$. Let S denote either A or B, but a fixed one of them on a given line. Trivially,

(26) $$aSb \text{ implies } a'Sb',$$

if $a' \equiv ha$, $b' \equiv hb$, $\pmod{4l}$, h an odd integer;

(27) $$aSb \text{ implies } bSa;$$

(28) $$aAa;$$

(29) $$aSb,\ bSc, \text{ imply } aAc;$$

(30) $$aAb,\ bBc, \text{ imply } aBc.$$

If $a'' = ia_1 - ja_2 - ka_3$, $-ia_1 + ja_2 - ka_3$, or $-ia_1 - ja_2 + ka_3$ is obtained from a by

(31) changing the signs of any two of the a_i,

then aAa'', and

(31′) $$aSb \text{ implies } a''Sb,\ aSb'',\ a''Sb''.$$

For $a'' = -iai$, $-jaj$, or $-kak$, and $N(ti) = Nt$, etc.

The third conclusion in (31′) may be generalized: *if the same sequence of permutations and sign-changes be applied to the coördinates of both a and b to produce $a^{\dagger}$ and $b^{\dagger}$, then*

(32) $$aSb \text{ implies } a^{\dagger}Sb^{\dagger}.$$

This is evident as regards an even number of permutations and sign-changes: for an even number of such operations on i, j, k produces a permutation of the entire class of integral quaternions which preserves the truth of all relations constructed by addition, multiplication, etc. If an odd number of operations are involved in forming $a^{\dagger}$ we note that aSb implies $(-a)S(-b)$, by (26), and use $-a^{\dagger}$ and $-b^{\dagger}$ in (32).

Examples: Write (x_1, x_2, x_3) for $ix_1 + jx_2 + kx_3$. From $(1,1,1)A(1,1,1)$ follows $(1,1,1)A(1,-1,-1)$, etc. From $(1,1,2)B(1,1,-2)$ follows by (26) $(3,3,-2)B(3,3,2)$, $(3,-3,-2)B(3,-3,2)$, etc.

10. The necessity of (10) was observed in §§ 3 and 4.

Both (10) and § 9, especially (32), reduce considerably the number of pairs of residue triples to be considered, first as to whether they can be transformed into each other, (mod 4) if of odd norm, (mod 8) if of norm $4n+2$; second, whether the transformation is of type A or B. All cases can be reduced to (33)-(38) below.

To begin with, suppose a_1, a_2, a_3 odd. We have

(33) $$(1,1,1)A(1,1,1), \qquad (1,1,1)B(-1,-1,-1).$$

For (33_2) follows from

$$(-i-j-k)(i+j+k)(i+j+k) \equiv -i-j-k \pmod{4}.$$

Changing any two signs at once we have the parts of Theorems 4 and 5 which concern $Na \equiv 3 \pmod{8}$.

Next let a_1 be odd, a_2 and a_3 even. We have

(34) $$(1,0,0)B(1,2,2), \qquad (1,2,0)B(1,0,2).$$

For

$$(1-i-j)i(1+i+j) = i+2j+2k,$$
$$(1-i-j)(i+2j)(1+i+j) = 5i+4j-2k \equiv i+2k \pmod{4}.$$

The parts of Theorems 4, 5 concerning $Na \equiv 1 \pmod{4}$ now follow.

We prove next the sequences of relations

(35) $$(1,1,0)A(3,3,4)B(1,1,4)A(3,3,0),$$

$$(36) \qquad (1,3,0)A(3,1,4)B(1,3,4)A(3,1,0):$$

$$(2i-k)(i+j)(2i+k)=3i-5j+4k\equiv 3i+3j+4k \pmod 8,$$
$$(2-i-j-k)(i+j)(2+i+j+k)=9i+j+4k\equiv i+j+4k,$$
$$(1-i-j)(i+j)(1+i+j)=3i+3j;$$
$$(-2i-k)(i+3j)(2i+k)=3i-15j+4k\equiv 3i+j+4k,$$
$$(i-j+k)(i+3j)(-i+j-k)=-7i-5j-4k\equiv i+3j+4k.$$

Hence, for example, $(1,1,0)B(1,1,4)$ by (30); none of the residue triples in (35) can be transformed into any in (36), in view of (10). Similarly

$$(37) \qquad (1,1,2)A(3,3,-2)B(1,1,-2)A(3,3,2),$$

$$(38) \qquad (1,3,2)A(3,1,-2)B(1,3,-2)A(3,1,2):$$

$$(-2i-k)(i+j+2k)(2i+k)=11i-5j-2k\equiv 3i+3j-2k,$$
$$(-i+j-k)(i+j+2k)(i-j+k)=i-7j-2k\equiv i+j-2k;$$
$$(-2i-k)(i+3j+2k)(2i+k)=11i-15j-2k\equiv 3i+j-2k,$$
$$(1+j+k)(i+3j+2k)(1-j-k)=(-3i+j-2k)A(3i+j+2k).$$

From (35)-(38), Theorems 4 and 5 may now be easily verified for $l=2$.

11. Part I of Theorem 1 will follow from the case $\lambda=1$ of

THEOREM 7. *Let a satisfy (2), (5), and (6), and let $a^\dagger$ be obtained from a by merely permuting or changing signs of a_1, a_2, a_3. Let λ be an odd integer prime to $h=Na$. Let H be any positive odd factor of h such that $h/(lH)\equiv 1 \pmod 8$, e. g. $H=h/l$. Then if the g. c. d. of the components of $\bar{t}at$ is prime to H, and t, u are of odd norms,*

$$(39) \qquad \bar{t}at\equiv\lambda\bar{u}a^\dagger u \pmod{4lH} \quad \textit{implies} \quad Nt\equiv(lH|\lambda)Nu \pmod 4.$$

We first extract from § 10 what pertains to $b\equiv\lambda a^\dagger \pmod{4l}$. In (39) necessarily $a^\dagger\equiv a \pmod 2$; and the number of values of $a^\dagger$ to be considered may be diminished by employing (31), which cannot affect (39), since $\bar{u}a^\dagger u=\bar{v}a^\ddagger v$, *if* $v=iu$ and $a^\ddagger=-ia^\dagger i$, etc. All cases are virtually contained then in

LEMMA 1. *If $Na\equiv 3 \pmod 8$, and $a'=a$ or is obtained from a by permuting the a_i, then aAb if $b\equiv a' \pmod 4$, aBb if $b\equiv -a' \pmod 4$. If $Na\not\equiv 3 \pmod 8$, let a' be obtained from a by interchanging the coördinates a_f and a_g of like parity; and a'' from a' by changing the sign of the remaining coördinate. Then:*

$$
\begin{array}{llll}
\textit{if } Na \equiv 1 \pmod{8}, & aAb \textit{ when } b \equiv \pm a,\ \pm a',\ \textit{or } \pm a'' & & \pmod{4}; \\
\textit{if } Na \equiv 5 \pmod{8}, & aAb \textit{ when } b \equiv \pm a & & \pmod{4}, \\
& aBb \textit{ when } b \equiv \pm a' \textit{ or } \pm a'' & & \pmod{4}; \\
\textit{if } Na \equiv 2 \pmod{16}, & aAb \textit{ when } b \equiv \pm a,\ \pm a',\ \textit{or } \pm a'' & & \pmod{8}, \\
& aBb \textit{ when } b \equiv \pm 3a,\ \pm 3a',\ \textit{or } \pm 3a'' & & \pmod{8}; \\
\textit{if } Na \equiv 6 \pmod{16}, & aAb \textit{ when } b \equiv a,\ 3a,\ a',\ 3a',\ 5a'',\ 7a'' & & \pmod{8}, \\
& aBb \textit{ when } b \equiv 5a,\ 7a,\ 5a',\ 7a',\ a'',\ 3a'' & & \pmod{8}; \\
\textit{if } Na \equiv 10 \pmod{16}, & aAb \textit{ when } b \equiv a,\ 7a,\ 3a',\ 5a',\ 3a'',\ 5a'' & & \pmod{8}, \\
& aBb \textit{ when } b \equiv 3a,\ 5a,\ a',\ 7a',\ a'',\ 7a'' & & \pmod{8}; \\
\textit{if } Na \equiv 14 \pmod{16}, & aAb \textit{ when } b \equiv a,\ 3a,\ 5a',\ 7a',\ a'',\ 3a'' & & \pmod{8}, \\
& aBb \textit{ when } b \equiv 5a,\ 7a,\ a',\ 3a',\ 5a'',\ 7a'' & & \pmod{8}.
\end{array}
$$

We obtain next certain relations connected with

$$(40)\qquad at = -x_0 + ix_1 + jx_2 + kx_3 = \xi, \quad \bar{t}at = ic_1 + jc_2 + kc_3 = c,$$

where, on expanding, we find

$$(41)\qquad \begin{array}{ll} x_0 = a_1t_1 + a_2t_2 + a_3t_3, & x_1 = a_1t_0 + a_2t_3 - a_3t_2, \\ x_2 = a_2t_0 + a_3t_1 - a_1t_3, & x_3 = a_3t_0 + a_1t_2 - a_2t_1; \end{array}$$

$$(42)\qquad \begin{array}{l} c_1 = a_1(t_0^2 + t_1^2 - t_2^2 + t_3^2) + 2a_2(t_0t_3 + t_1t_2) + 2a_3(-t_0t_2 + t_1t_3), \\ c_2 = 2a_1(-t_0t_3 + t_1t_2) + a_2(t_0^2 + t_2^2 - t_1^2 - t_3^2) + 2a_3(t_0t_1 + t_2t_3), \\ c_3 = 2a_1(t_0t_2 + t_1t_3) + 2a_2(-t_0t_1 + t_2t_3) + a_3(t_0^2 + t_3^2 - t_1^2 - t_2^2). \end{array}$$

We can easily verify that, to modulus h $(= \Sigma a_i^2)$,

$$(43)\qquad \begin{array}{lll} a_1c_1 \equiv x_0^2 + x_1^2, & a_1c_2 \equiv x_0x_3 + x_1x_2, & a_1c_3 \equiv -x_0x_2 + x_1x_3, \\ a_2c_1 \equiv -x_0x_3 + x_1x_2, & a_2c_2 \equiv x_0^2 + x_2^2, & a_2c_3 \equiv x_0x_1 + x_2x_3, \\ a_3c_1 \equiv x_0x_2 + x_1x_3, & a_3c_2 \equiv -x_0x_1 + x_2x_3, & a_3c_3 \equiv x_0^2 + x_3^2; \end{array}$$

$$(44)\qquad \Sigma x_f^2 = N\xi = Na \cdot Nt \equiv 0 \pmod{h}.$$

Similarly we write

$$(45)\qquad a^\dagger u = -y_0 + iy_1 + jy_2 + ky_3 = \eta, \qquad \bar{u}a^\dagger u = id_1 + jd_2 + kd_3 = d.$$

The substitutions which carry a into $a^\dagger$ replace (41)-(44) by ($41^\dagger$)-($44^\dagger$) with $t \to u$, $x \to y$, $c \to d$, $\xi \to \eta$.

In order that $\bar{t}at$ be prime to H, at must be likewise, i. e.

$$(46)\qquad x_0, x_1, x_2, x_3, H \text{ must be coprime.}$$

The hypothesis of (39) breaks up into two parts:

$$(47) \qquad c_i \equiv \lambda d_i \pmod{H} \qquad (i = 1, 2, 3),$$

$$(47') \qquad \bar{u}\bar{a}t\bar{a}\bar{u} \equiv \lambda a^\dagger \pmod{4l}.$$

We shall show in all cases that, if (47) and (46) both hold, one of the four characters $(\pm \lambda|H)$, $(\pm 2\lambda|H)$ has a certain value, and then verify from Lemma 1 that in accordance with this value and (47′),

$$(48) \qquad aA\lambda a^\dagger \text{ if } (lH|\lambda) = 1, \qquad aB\lambda a^\dagger \text{ if } (lH|\lambda) = -1.$$

First take $a^\dagger = a$. Now $a_i c_i \equiv \lambda a_i d_i \pmod{H}$, and by (43), (43†),

$$(49) \qquad x_0^2 + x_i^2 \equiv \lambda(y_0^2 + y_i^2) \pmod{H} \qquad (i = 1, 2, 3).$$

Summing, we have $2x_0^2 + N\xi \equiv \lambda(2y_0^2 + N\eta)$, $x_0^2 \equiv \lambda y_0^2$,

$$(50) \qquad x_f^2 \equiv \lambda y_f^2 \pmod{H} \qquad (f = 0, 1, 2, 3).$$

This contradicts (46) unless $(\lambda|H) = 1$, whence

$$(51) \qquad \begin{aligned}(lH|\lambda) &= -(l|\lambda) \text{ if } H \equiv \lambda \equiv 3 \pmod 4,\\ &= (l|\lambda) \text{ otherwise.}\end{aligned}$$

Now $H \equiv h/l \pmod 8$. In Lemma 1, if $h \equiv 3 \pmod 8$, $aA\lambda a$ or $aB\lambda a$ according as $\lambda \equiv 1$ or $-1 \pmod 4$, in agreement with (51); if $H \equiv 3 \pmod 4$ and $l = 2$, $aA\lambda a$ if $\lambda \equiv 1$ or $3 \pmod 8$, $aB\lambda a$ if $\lambda \equiv 5$ or $7 \pmod 8$, which agrees with (51). If $H \not\equiv 3 \pmod 4$ a glance at Lemma 1 confirms that $aA\lambda a$ or $aB\lambda a$ according as $(l|\lambda) = 1$ or -1.

Second take $a^\dagger = ia_1 + ja_3 + ka_2$, assuming $a_2 \equiv a_3 \pmod 2$. This is the a' of Lemma 1. The cases $a_1 \equiv a_2$ or $a_3 \equiv a_1 \pmod 2$ will follow by symmetry.

Now by (43) and (43†), and (44†),

$$\begin{gathered}a_1c_1 \equiv -x_2^2 - x_3^2,\ a_2c_2 + a_3c_3 \equiv 2x_0^2 + x_2^2 + x_3^2,\ a_2c_3 + a_3c_2 \equiv 2x_2x_3,\\ a_1d_1 \equiv -y_2^2 - y_3^2,\ a_3d_3 + a_2d_2 \equiv 2y_2y_3,\ a_3d_2 + a_2d_3 \equiv 2y_0^2 + y_2^2 + y_3^2,\end{gathered}$$

and by (47),

$$(52) \qquad \begin{aligned}&2x_0^2 \equiv -\lambda(y_2 - y_3)^2, && 2\lambda y_0^2 \equiv -(x_2 - x_3)^2,\\ &(x_2 + x_3)^2 \equiv -2\lambda y_1^2, && \lambda(y_2 + y_3)^2 \equiv -2x_1^2, \pmod{H},\end{aligned}$$

whence, unless $(-2\lambda|H) = 1$, some prime factor of H divides x_0, x_1, $x_2 + x_3$, and $x_2 - x_3$, contrary to (46). Hence

$$(52') \qquad \begin{aligned}(lH|\lambda) &= -(l|\lambda)(-2|H) \text{ if } H \equiv \lambda \equiv 3 \pmod 4,\\ &= (l|\lambda)(-2|H) \text{ otherwise.}\end{aligned}$$

The comparison with (48) and Lemma 1 is left to the reader.

Third take $a^\dagger = -ia_1 + ja_3 + ka_2$, $a_2 \equiv a_3 \pmod{2}$. This is equivalent, by (31), to $a^\dagger = -ia_1 - ja_3 - ka_2$, and hence may be reduced to the preceding case by changing λ to $-\lambda$. Thus we have to compare (48) with $a^\dagger = a''$, and Lemma 1, upon the assumption that $(2\lambda|H) = 1$, i. e.

$$\begin{aligned}(lH|\lambda) &= -(l|\lambda)(2|H) \text{ if } H \equiv \lambda \equiv 3 \pmod{4},\\ &= \quad (l|\lambda)(2|H) \text{ otherwise.}\end{aligned}$$

The accordance with Lemma 1 is easily verified.

There now remains only $a_1 \equiv a_2 \equiv a_3 \pmod{2}$, whence $Na \equiv 3 \pmod{8}$; with $a^\dagger$ equal to α) $ia_2 + ja_3 + ka_1$, β) $-ia_2 - ja_3 - ka_1$, γ) $ia_3 + ja_1 + ka_2$, or δ) $-ia_3 - ja_1 - ka_2$.

Fourth, assume α). By (43) and (43†),

$$\begin{gathered}a_1c_1 \equiv x_0^2 + x_1^2 \equiv -x_2^2 - x_3^2,\ a_2c_3 \equiv x_0x_1 + x_2x_3,\ a_3c_2 \equiv -x_0x_1 + x_2x_3,\\ a_1d_1 \equiv y_0y_2 + y_1y_3,\ a_2d_3 \equiv -y_0y_2 + y_1y_3,\ a_3d_2 \equiv y_0^2 + y_2^2 \equiv -y_1^2 - y_3^2,\end{gathered}$$

whence by (47),

$$\begin{aligned}(x_0 + x_1)^2 &\equiv \lambda(y_1 + y_3)^2, & (x_0 - x_1)^2 &\equiv \lambda(y_0 + y_2)^2,\\ (x_2 + x_3)^2 &\equiv \lambda(y_0 - y_2)^2, & (x_2 - x_3)^2 &\equiv \lambda(y_1 - y_3)^2,\end{aligned}$$

contradicting (46) unless $(\lambda|H) = 1$. That is, since $Na = h \equiv H \equiv 3 \pmod{8}$, $(H|\lambda) = (-1|\lambda)$. This agrees with the first part of Lemma 1 and (48).

Case β) is reduced to α) by changing λ to $-\lambda$; hence $(-\lambda|H) = 1$ and $(H|\lambda) = -(-1|\lambda)$, in agreement with (48) and Lemma 1 with $Na \equiv 3 \pmod{8}$. Cases γ) and δ) are similar.

As a corollary we have

THEOREM 8. *Let H be an odd factor of Na, a as in (5), λ an integer. Then the congruence*

(53) $$\bar{t}at \equiv \lambda\bar{u}a^\dagger u \pmod{H},$$

has no solutions t, u with $\bar{t}at$ proper, if

1) *λ is a quadratic non-residue of H, and (apart from operations*
(53′) *(31)), $a^\dagger = a$ or $ia_2 + ja_3 + ka_1$ or $ia_3 + ja_1 + ka_2$;*

or

2) *-2λ is a quadratic non-residue of H, and (apart from operations*
(53″) *(31)), $a^\dagger = ia_1 + ja_3 + ka_2$ or $ia_3 + ja_2 + ka_1$ or $ia_2 + ja_1 + ka_3$.*

By changing λ into $-\lambda$, $a^\dagger$ into $-a^\dagger$, we have a result associated with any odd or even number of permutations or of sign-changes of the a_i.

THEOREM 9. *Let H be any odd integer, λ an integer prime to H, a proper*

and pure. Then (53) *is solvable for* t, u *with* $\bar{t}at$ *proper, if* (53′) *holds and* λ *is a quadratic residue of* H, *or if* (53″) *holds and* -2λ *is a quadratic residue of* H.

For let u be any quaternion for which $\bar{u}a^{\dagger}u$ is proper (mod H), e. g. $u=1$. If $\lambda\equiv s^2$ (mod H), and $a^{\dagger}=a$, take $t\equiv su$; if however $a^{\dagger}=ia_2+ja_3+ka_1$, take $t\equiv su(1+i+j+k)/2$ (mod H); and if $a^{\dagger}=ia_3+ja_1+ka_2$, take $t\equiv su(1-i-j-k)/2$. The trivial effect of operations (31) is by now obvious. If $\lambda\equiv-2s^2$ (mod H), and $a^{\dagger}=ia_1-ja_3-ka_2$, take $t\equiv s(j+k)$; and so forth.

Taking $\lambda=1$ we have at once the statement referred to § 11 on p. 895.

12. THEOREM 10. *Let* a, b *be pure integral quaternions,*

$$p \text{ an odd prime, } p\nmid a,\ p\nmid b,\ p^r\,|\,Na,\ p^r\,|\,Nb,\ r>0. \tag{54}$$

Then the number of solutions t (mod p^r) *of*

$$\bar{t}at\equiv b \pmod{p^r} \tag{55}$$

is exactly p^{2r-2} *times the number of solutions* t (mod p) of

$$\bar{t}at\equiv b \pmod{p}. \tag{56}$$

Clearly, any solution t of (55) must satisfy

$$\bar{t}at\equiv b \pmod{p^{r-1}}. \tag{57}$$

The solutions of (55) are to be sought among the quaternions $t+up^{r-1}$, where t ranges over all solutions of (57), and u is to be determined (mod p) to satisfy

$$(\bar{t}+p^{r-1}\bar{u})a(t+p^{r-1}u)\equiv b \pmod{p^r}.$$

On setting $\bar{t}at-b=p^{r-1}v$, this reduces to

$$v+\bar{u}at-\bar{t}au\equiv 0 \pmod{p},$$

and hence on writing $at=-x_0+\cdots$ as in (40), we have

$$\begin{aligned} x_1u_0+x_0u_1-x_3u_2+x_2u_3&\equiv-\tfrac{1}{2}v_1,\\ x_2u_0+x_3u_1+x_0u_2-x_1u_3&\equiv-\tfrac{1}{2}v_2,\\ x_3u_0-x_2u_1+x_1u_2+x_0u_3&\equiv-\tfrac{1}{2}v_3, \pmod{p}. \end{aligned} \tag{58}$$

These three congruences are not independent. For if $c=\bar{t}at$,

$$\xi c=at\bar{t}at=-Nt(Na)t\equiv 0,$$

whence

$$(59)\quad \begin{aligned} &c_1x_1 + c_2x_2 + c_3x_3 \equiv 0, \qquad &&c_1x_0 + c_2x_3 - c_3x_2 \equiv 0,\\ -&c_1x_3 + c_2x_0 + c_3x_1 \equiv 0, &&c_1x_2 - c_2x_1 + c_3x_0 \equiv 0, \pmod{p}. \end{aligned}$$

Fortunately, the same relation connects the right members of (58):

$$(60)\qquad c_1v_1 + c_2v_2 + c_3v_3 \equiv 0 \pmod{p}.$$

To prove this we consider $c - b = p^{r-1}v$,

$$(\bar{c} - \bar{b})(c - b) = p^{2r-2}\bar{v}v,$$

whence as $p^r | Nc$ and $p^r | Nb$, $p^r | \frac{1}{2}(\bar{c}b + \bar{b}c) =$ real part of $\bar{c}b$. But $\bar{c}c - \bar{c}b = p^{r-1}\bar{c}v$. Hence p divides the real part of cv, which is (60).

Thus the congruences (58) reduce to two. For example let $p \nmid c_3$. Then (58_3) is a combination of (58_1) and (58_2). By (43), one of the determinants $-x_0x_2 + x_1x_3$, $x_0x_1 + x_2x_3$, $x_0^2 + x_3^2$ is prime to p. Hence we can choose two of u_0, u_1, u_2, u_3 arbitrary $\pmod{p}$, and the remaining two are then uniquely determined. That is, (58) has precisely p^2 solutions $u \pmod{p}$, and (55) has exactly p^2 times the number of solutions of (57).

13. THEOREM 11. *Let a be a pure quaternion, $p | Na$, $p \nmid a$, p an odd prime. Then*

$$(61)\qquad \bar{t}at \equiv 0 \pmod{p} \textit{ for exactly } p^2 \textit{ residues } t \pmod{p};$$

and $\bar{t}at$ represents $\pmod{p}$ exactly $\frac{1}{2}(p^2 - 1)$ of the $p^2 - 1$ residues b such that $p \nmid b$ and $p | Nb$, and each such residue represented is obtained for exactly $2p^2$ incongruent values $t \pmod{p}$.

We inquire, for a given u, how many values $t \pmod{p}$ satisfy

$$(62)\qquad \bar{t}at \equiv \bar{u}au \pmod{p}.$$

We employ notations (40)-(45), with $a^\dagger = a$. Hence (62) becomes $c_i \equiv d_i \pmod{p}$ $(i = 1, 2, 3)$.

Thus (62) implies $a_ic_i \equiv a_id_i$, and as in (50),

$$(63)\qquad x_f^2 \equiv y_f^2 \pmod{p}, \qquad (f = 0, 1, 2, 3).$$

Conversely either of the systems

$$(64)\qquad x_f \equiv y_f \pmod{p}, \qquad (f = 0, 1, 2, 3),$$

$$(65)\qquad x_f \equiv -y_f \pmod{p}, \qquad (f = 0, 1, 2, 3),$$

implies in view of (43) and $(43^\dagger)$, $a_ic_j \equiv a_id_j$ $(i, j = 1, 2, 3)$, and hence, since $p \nmid a$, $c_j \equiv d_j$ $(j = 1, 2, 3)$, which is (62).

Now (64) and (65) each possess exactly p^2 incongruent solutions $t \pmod p$. To see this we write out (64) in full:

$$(66)\quad \begin{array}{llll} & a_1t_1 + a_2t_2 + a_3t_3 \equiv & & a_1u_1 + a_2u_2 + a_3u_3, \\ a_1t_0 & \quad - a_3t_2 + a_2t_3 \equiv a_1u_0 & & - a_3u_2 + a_2u_3, \\ a_2t_0 + a_3t_1 & \quad - a_1t_3 \equiv a_2u_0 + a_3u_1 & & - a_1u_3, \\ a_3t_0 - a_2t_1 + a_1t_2 & \quad \equiv a_3u_0 - a_2u_1 + a_1u_2, & & \pmod p. \end{array}$$

The argument for $p \nmid a_1$ will be typical. Then (66_3), (66_4) are linear combinations of (66_1) and (66_2), with the multipliers a_3/a_1 and a_2/a_1, $-a_2/a_1$ and a_3/a_1. The system reduces to the first two congruences; t_2 and t_3 can be chosen arbitrarily, and t_0, t_1 are then uniquely determined; that is, t has p^2 residues $(\bmod p)$. The same argument applies to (65).

The solutions t of (64) are distinct from those of (65) unless $y_f \equiv -y_f \equiv 0$ $(f = 0, 1, 2, 3)$. Then $au \equiv 0$, $\bar{u}au \equiv 0$. The case $\bar{t}at \equiv 0 \pmod p$ can now be completed. We can take $u = 0$, and see that (66) with zeros on the right has again exactly p^2 solutions. This yields (61). Incidentally it shows that

$$(67)\qquad \text{if } p \nmid a,\ p \mid Na, \text{ then } p \mid \bar{t}at \text{ if and only if } p \mid at.$$

For the rest we shall suppose $p \nmid c = \bar{t}at$. In addition to (59) we must, if $c_i \equiv d_i$, have

$$(68)\quad \begin{array}{ll} c_1y_1 + c_2y_2 + c_3y_3 \equiv 0, & c_1y_0 + c_2y_3 - c_3y_2 \equiv 0, \\ -c_1y_3 + c_2y_0 + c_3y_1 \equiv 0, & c_1y_2 - c_2y_1 + c_3y_0 \equiv 0, \pmod p. \end{array}$$

There remains from (63) to be treated the possibilities $x_f \equiv \pm y_f$ in which not all the signs are alike. We shall reduce all cases to (64) or (65).

Suppose for example, $x_0 \equiv -y_0$, $x_1 \equiv y_1$, $x_2 \equiv y_2$, $x_3 \equiv y_3$. From (59) and (68) follow $p \mid 2c_iy_0$ $(i = 1, 2, 3)$, $p \mid y_0$. Hence $x_0 \equiv y_0$ and we have (64). Similarly whenever one or three of the signs are $+$ (or $-$) we are led to (64) or (65).

The case of two plus or two minus signs is typified by

$$(69)\qquad x_0 \equiv -y_0,\ x_1 \equiv -y_1,\ x_2 \equiv +y_2,\ x_3 \equiv +y_3.$$

If $c_i \equiv d_i$, then (59) and (68) yield

$$p \mid c_1,\ p \mid c_2y_2 + c_3y_3,\ p \mid c_2y_0 + c_3y_1,$$

and since $p \nmid c$, $p \mid c_2^2 + c_3^2$ but $p \nmid c_2c_3$. Incidentally, $p \equiv 1 \pmod 4$. Further by (43),

$$a_1c_2 \equiv x_0x_3 + x_1x_2 \equiv -y_0y_3 - y_1y_2 \equiv -a_1d_2,$$

whence $p|a_1$ also (since $c_2 \equiv d_2 \not\equiv 0$), and $p \nmid a_2a_3$. Thus

$$x_0 \equiv a_2t_2 + a_3t_3 \equiv (-a_2/a_3)(-a_3t_2 + a_2t_3) \equiv (-a_2/a_3)x_1,$$

and similarly $x_2 \equiv (a_2/a_3)x_3$. Thus by (42_1) and $p|c_1$,

$$\begin{aligned} 0 \equiv c_1 &\equiv 2t_0(a_2t_3 - a_3t_2) + 2t_1(a_2t_2 + a_3t_3) \\ &\equiv 2t_0x_1 - 2t_1(a_2/a_3)x_1 \equiv (2/a_3)x_1(a_3t_0 - a_2t_1) \equiv 2x_1x_3/a_3, \end{aligned}$$

that is either $x_0 \equiv x_1 \equiv 0$, or $x_2 \equiv x_3 \equiv 0$; reducing the case to (64) or (65).

The theorem follows, since t has p^4 possible residues (mod p), p^2 of these are used for $\bar{t}at \equiv 0$, and the remaining $p^2(p^2-1)$ form sets of $2p^2$ residues t for each of which $\bar{t}at$ has the same residue. That

$$(70) \qquad b_1^2 + b_2^2 + b_3^2 \equiv 0 \pmod{p}$$

has exactly p^2 solutions, including $(0, 0, 0)$, is easy to show.

The characterization by other criteria of the residues $\bar{t}at \pmod{p}$ is not here attempted.

COROLLARY 1. *Congruence* (56) *is solvable for* $b \equiv 0$ *and for only half those residues* $b \not\equiv 0 \pmod{p}$ *such that* $Nb \equiv 0$.

COROLLARY 2. *If* $p|Na$, $p > 2$, *then* $\bar{t}at \equiv \bar{u}au$ *implies* $at \equiv au$ *or* $at \equiv -au \pmod{p}$.

14. Write $a^* = ia_1 + ja_3 + ka_2$, and assume $gcd(a_1, a_2, a_3) = 1$. By Theorem 8 with $\lambda = \pm 1$, and Theorem 11 *et seq.*, we see that if H is an odd factor of Na, then all residues $b \pmod{H}$, such that $Nb \equiv 0$, are represented between

$$(71) \quad \begin{array}{lll} \bar{t}at \text{ and } \bar{t}(-a)t, & \text{when } H = p \equiv 3 \pmod{4}, \\ \bar{t}at \text{ and } \bar{t}a^*t, & \text{when } H = p \equiv 5 \text{ or } 7 \pmod{8}, \\ \bar{t}at \text{ and } \bar{t}(-a^*)t, & \text{when } H = p \equiv 3 \text{ or } 5 \pmod{8}, \\ \pm\bar{t}at \text{ and } \pm\bar{t}a^*t, & \text{when } H = pq,\ p,\ q, \text{ and } 1 \text{ incongruent } \pmod{8}; \end{array}$$

p and q representing odd primes.

It will be observed that if, in Theorem 1, h' has no odd factor > 1 of the form $8f + 1$, it is of the form lH, with

$$(72) \qquad H = 1 \text{ or } p \text{ or } pq,$$

p, q odd primes incongruent to each other and to 1 (mod 8).

Evidently all solutions of (1) satisfy

$$(73)\qquad Nx \equiv lH \pmod{8lH},$$

if H is an odd divisor of h such that $h/(lH) \equiv 1 \pmod 8$. Further, if H is of type (72), then as a ranges over a set $\Sigma(a)$, and t over all integral quaternions, $\bar{t}at$ represents *all* residues $x \pmod{4lH}$ satisfying (73). For by Theorem 4, since $Na = h \equiv lH \pmod{8l}$, every residue $x \pmod{4l}$ such that

$$(74)\qquad x \equiv a \pmod 2, \qquad Nx \equiv lH \pmod{8l}$$

is represented by each of $\bar{t}at$ and $-\bar{t}at$ (separately), and every residue $x \equiv a^* \pmod 2$ satisfying (74_2) by each of $\bar{t}a^*t$ and $-\bar{t}a^*t$. We can choose a so that $a_2 \equiv a_3 \pmod 2$, $a \equiv a^* \pmod 2$, and hence can combine all residues $\pmod H$ with all residues $\pmod{4l}$ such that $x \equiv a \pmod 2$. Permuting a_1, a_2, a_3 cyclically we get all residues $x \pmod{4lH}$ subject to (73). This proves the first part of II in Theorem 1.

15. To prove the second part of II in Theorem 1, consider a solution x of (1) with $(h/h')(8n+1)$ not a square. We can choose an odd prime p such that simultaneously

$$(75)\qquad (-h'|p) = -1, \qquad (-h(8n+1)|p) = 1.$$

Choose s so that $s^2 + h(8n+1) \equiv 0 \pmod p$. Then $s + x$ has exactly eight (left-associate) right divisors of norm p, say

$$(76)\qquad s + x = uv, \qquad Nv = p.$$

Then

$$(77)\qquad vx\bar{v} = px', \text{ where } x' = vu - s.$$

The relation between x and x' is reciprocal, $\bar{v}$ carrying x' back into x. If v is replaced by a left-associate, the only change in x' is that two components are changed in sign. Thus a $(1,1)$ correspondence is set up connecting the four solutions

$$(x_1, x_2, x_3),\ (x_1, -x_2, -x_3),\ (-x_1, x_2, -x_3),\ (-x_1, -x_2, x_3),$$

of (1), with four similarly related solutions x', $-ix'i$, etc.

We shall prove that x and x' are in opposite classes A and B, in view of (75_1), provided $h' = lH$, H of form (72). Suppose $x \equiv \bar{t}at \pmod{4h'}$, which is possible by § 14; and let $x' \equiv \bar{w}a^\dagger w \pmod{4h'}$, where $a^\dagger$ is also in $\Sigma(a)$. Then by (77),

$$v\bar{t}at\bar{v} \equiv p\bar{w}a^{\dagger}w \pmod{4h'}.$$

By Theorem 7,

$$N(vt) \equiv Nw \pmod 4 \quad \text{if} \quad (h'|p) = 1,$$
$$\equiv -Nw \qquad \text{if} \quad (h'|p) = -1.$$

If $(h'|p) = 1$, $p \equiv 3 \pmod 4$ by (75_1), $Nt \equiv -Nw \pmod 4$, x' is in the opposite class to x. If $(h'|p) = -1$, $p \equiv 1 \pmod 4$, and again $Nv \equiv -Nw$.

16. In Theorem 1 let $h' = lH$ have no factor > 1 and $\equiv 1 \pmod 8$. By § 14 every solution x of (73) with x_1, x_2, x_3, h' coprime, belongs to A or B. By Theorem 7,

(78) if λ is an integer prime to $4h'$, then x and λx are in the same or opposite classes A or B, according as $(h'|\lambda) = 1$ or -1.

For example, in the table following Theorem 1, if $h' = 2$ and $\lambda = 3$, $(2|3) = -1$, $(1, 1, 0)$ is in A, $(3, 3, 0)$ in B.

17. If $h = H$ or $2H$, H as in (72), then for any solution x of (1) there exist integers r_1, r_2, r_3, and an odd integer m such that

$$hm^2 = (x_1 + 4r_1h)^2 + (x_2 + 4r_2h)^2 + (x_3 + 4r_3h)^2. \tag{79}$$

For we can solve $x \equiv \bar{t}at \pmod{4h}$ with $Na = h$ and Nt odd.

McGill University.

A LIST OF EXPANSIONS FOR THE FUNCTION $\phi(x, y, z)$.*

By Walter H. Gage.

1. In a recent paper [1] the writer has given a method for obtaining the trigonometric expansions for the so-called doubly periodic functions $\phi_{abcd}(x,y,z)$, where

$$\phi_{abcd}(x, y, z) \equiv \frac{\vartheta'_1{}^2 \vartheta_a(x+y+z)}{\vartheta_b(x)\vartheta_c(y)\vartheta_d(z)},$$

for sixty-four values of the quadruple index $abcd$.

Using the set of sixteen expansions for the so-called doubly periodic functions $\phi_{abc}(x, y)$ of the second kind [2] (first obtained by Hermite) and the relation

$$\phi_{abcd}(x, y, z) = \phi_{ebc}(x, y) \cdot \phi_{aed}(x+y, z),$$

we obtain expansions which are valid where the regions of convergence of the functions $\phi_{ebc}(x, y)$, $\phi_{aed}(x+y, z)$ overlap.[3] By this method it is possible to get the expansions for $abcd$ equal to 0000, 1111, 2222, 3333, 0011, 0022, 0033, 1122, 1133, 2233, 0123, and all possible arrangements of each of these numbers. In the lists below, however, we have omitted those which are obtainable from the given ones merely by interchanging x, y, z.

The two sets of expansions obtained by Basoco and Bell [4] for forty-eight functions in two variables prove to be special cases of these new expansions for x or y equal to zero. The case $z = 0$ gives us for the forty-eight functions a third set which so far has not been published. On the other hand, the method employed by Basoco and Bell can be applied to the sixty-four $\phi_{abcd}(x, y, z)$ to yield four sets of expansions for one hundred and ninety-two additional functions in three variables.

* Received June 17, 1937.

[1] "A method of obtaining certain theta expansions," to appear in the *Transactions of the Royal Society of Canada*.

[2] E. T. Bell, "Arithmetical paraphrases," *Transactions of the American Mathematical Society*, vol. 22 (1921), pp. 198-219; "Theta expansions useful in arithmetic," *Messenger of Mathematics*, no. 635, vol. 54 (1924), pp. 166-176.

[3] Tannery and Molk, "Éléments de la Théorie des Fonctions Elliptiques," Tome 3, pp. 120-129.

[4] "Further theta expansions useful in arithmetic," *American Journal of Mathematics*, vol. 57 (1935), pp. 1-10.

2. As in the case of the series in two variables, these new expansions enable us to derive by the method of paraphrase a large body of arithmetical theorems on quadratic forms. In particular a full set, applied systematically to Jacobi's identities [5] (and to all which can be derived from them) expressing $\vartheta_\alpha(x+y+z)$ in terms of $\vartheta_\beta(x)$, $\vartheta_\gamma(y)$, $\vartheta_\delta(z)$, leads to an endless variety of paraphrases, involving arbitrary odd or even functions of divisors of integers in given partitions. In the general case, relations between ϑ and ϕ functions enable us to get paraphrases in any even number of variables. As a matter of fact, since certain of the $\phi_{abcd}(x, y, z)$ possess alternate expansions, direct comparison of the two series in these cases gives at once a set of arithmetical theorems. A study of the results that have been obtained by using the expansions for $\phi_{abc}(x, y)$ gives ample evidence of the possible applications of these new expansions in deriving theorems on quadratic partitions.

In what follows the reader is referred to Bell's papers cited above,[2] and in particular to the paper by Basoco and Bell,[4] for explanations of notation.

3. Following their method we set down the various partitions which occur.

(I) $\quad 2n = 2n_1 + 2n_2, \quad n = d\delta, \quad n_1 = d_1\delta_1, \ n_2 = d_2\delta_2.$

(II) $\quad n = n_1 + n_2, \quad n = t\tau, \quad n_1 = t_1\tau_1, \ n_2 = t_2\tau_2.$

(III) $\quad m = 4n_1 + m_2, \quad m = t\tau, \quad n_1 = d_1\delta_1, \ m_2 = t_2\tau_2.$

(IV) $\quad \begin{cases} m = n_1 + 2n_2, \quad m = t\tau, \quad n_1 = t_1\tau_1, \ n_2 = d_2\delta_2. \\ 2n = n'_1 + 2n'_2, \ 2n = t'\tau', \ n = d\delta, \quad n'_1 = t'_1\tau'_1, \ n'_2 = d'_2\delta'_2. \end{cases}$

(V) $\quad m = m_1 + 2n_2, \quad m = t\tau, \quad m_1 = t_1\tau_1, \quad n_2 = t_2\tau_2.$

4. The expansions of which the partition is (I) (in § 3) have the form

$$\phi_{abcd}(x, y, z) = T_{abcd}(x, y, z) + 4 \sum_{(n)} q^{2n} (\sum F_{abcd}(x, y, z; n)).$$

If we let

$$P_1(x, y, z) = \cos 2\{(d_1 - d_2)x + (d_1 - \delta_2)y + \delta_1 z\} \\ - \cos 2\{(d_1 + d_2)x + (d_1 + \delta_2)y + \delta_1 z\}$$

then

$T_{1111}(x, y, z) = -1 + \cot x \cot y + \cot y \cot z + \cot z \cot x;$
$T_{2211}(x, y, z) = -1 - \tan x \cot y + \cot y \cot z - \cot z \tan x;$
$T_{1221}(x, y, z) = 1 - \tan x \tan y + \tan y \cot z + \cot z \tan x;$
$T_{2112}(x, y, z) = T_{2211}(z, y, x); \quad T_{1122}(x, y, z) = T_{1221}(z, y, x);$
$T_{2222}(x, y, z) = 1 - \tan x \tan y - \tan y \tan z - \tan z \tan x;$

[5] Jacobi, *Werke*, vol. 1, p. 501 etc.

$$
\begin{aligned}
F_{1111}(x, y, z;\, n) = F_{1111}(x, y, z) &= \{\cot(x+y) + \cot z\} \sin 2(dx + \delta y) \\
&\quad + (\cot x + \cot y) \sin 2\{d(x+y) + \delta z\} + 2P_1(x, y, z); \\
F_{2211}(x, y, z) &= (-1)^d[\{\cot z - \tan(x+y)\} \sin 2(dx + \delta y) \\
&\quad + (\cot y - \tan x) \sin 2\{d(x+y) + \delta z\}] \\
&\quad + 2(-1)^{d_1+d_2} P_1(x, y, z); \\
F_{1221}(x, y, z) &= -(-1)^{d+\delta}\{\cot z + \cot(x+y)\} \sin 2(dx + \delta y) \\
&\quad + (\tan x + \tan y) \sin 2\{d(x+y) + \delta z\} \\
&\quad - 2(-1)^{d_2+\delta_2} P_1(x, y, z); \\
F_{2112}(x, y, z) &= \{\cot(x+y) - \tan z\} \sin 2(dx + \delta y) \\
&\quad + (-1)^\delta (\cot x + \cot y) \sin 2\{d(x+y) + \delta z\} \\
&\quad + 2(-1)^{\delta_1} P_1(x, y, z); \\
F_{1122}(x, y, z) &= (-1)^\delta[\{\tan(x+y) + \tan z\} \sin 2(dx + \delta y) \\
&\quad - (-1)^d (\cot x - \tan y) \sin 2\{d(x+y) + \delta z\}] \\
&\quad - 2(-1)^{d_1+\delta_1+\delta_2} P_1(x, y, z); \\
F_{2222}(x, y, z) &= (-1)^\delta[(-1)^d\{\tan z - \cot(x+y)\} \sin 2(dx + \delta y) \\
&\quad + (\tan x + \tan y) \sin 2\{d(x+y) + \delta z\}] \\
&\quad - 2(-1)^{d_2+\delta_1+\delta_2} P_1(x, y, z).
\end{aligned}
$$

5. The expansions in which the partition is (II) have the form

$$\phi_{abcd}(x, y, z) = T_{abcd}(x, y, z) + 4 \sum_{(n)} q^n \left(\sum F_{abcd}(x, y, z;\, n)\right).$$

Let

$$
\begin{aligned}
P_2(x, y, z) &= H_1(x, y, z), \quad R_2(x, y, z) = H_2(x, y, z), \\
Q_2(x, y, z) &= K_1(x, y, z), \quad S_2(x, y, z) = K_2(x, y, z), \\
H_k(x, y, z) &= \cos\{2(t_1 - t_2)x + (2t_1 - \tau_2)y + \tau_1 z\} \\
&\qquad + (-1)^k \cos\{2(t_1 + t_2)x + (2t_1 + \tau_2)y + \tau_1 z\}, \\
K_k(x, y, z) &= \sin\{2(t_1 + t_2)x + (2t_1 + \tau_2)y + \tau_1 z\} \\
&\qquad - (-1)^k \sin\{2(t_1 - t_2)x + (2t_1 - \tau_2)y + \tau_1 z\}.
\end{aligned}
$$

We have

$$
\begin{aligned}
T_{0011}(x, y, z) &= T_{3311}(x, y, z) = \csc y \csc z; \\
T_{3021}(x, y, z) &= T_{0321}(x, y, z) = T_{3012}(x, z, y) \\
&= T_{0312}(x, z, y) = \sec y \csc z; \\
T_{0022}(x, y, z) &= T_{3322}(x, y, z) = \sec y \sec z; \\
F_{0011}(x, y, z) &= \csc y \sin\{2t(x+y) + \tau z\} \\
&\quad + \csc z \sin(2tx + \tau y) + 2P_2(x, y, z); \\
F_{3311}(x, y, z) &= (-1)^n[\csc y \sin\{2t(x+y) + \tau z\} \\
&\quad + \csc z \sin(2tx + \tau y) + 2P_2(x, y, z)]; \\
F_{3021}(x, y, z) &= (-1)^n \sec y \sin\{2t(x+y) + \tau z\} \\
&\quad + (-1 \,|\, \tau) \csc z \cos(2tx + \tau y) \\
&\quad + 2(-1)^{n_1}(-1 \,|\, \tau_2) Q_2(x, y, z);
\end{aligned}
$$

$$
\begin{aligned}
F_{0321}(x, y, z) &= \sec y \sin\{2t(x+y)+\tau z\} \\
&\quad + (-1)^{n}(-1\,|\,\tau)\csc z \cos(2tx+\tau y) \\
&\quad + 2(-1)^{n_2}(-1\,|\,\tau_2) Q_2(x, y, z)\,; \\
F_{3012}(x, y, z) &= (-1\,|\,\tau)\csc y \cos\{2t(x+y)+\tau z\} \\
&\quad + \sec z \sin(2tx+\tau y) + 2(-1\,|\,\tau_1) S_2(x, y, z)\,; \\
F_{0312}(x, y, z) &= (-1)^{n}[(-1\,|\,\tau)\csc y \cos\{2t(x+y)+\tau z\} \\
&\quad + \sec z \sin(2tx+\tau y) + 2(-1\,|\,\tau_1) S_2(x, y, z)]\,; \\
F_{0022}(x, y, z) &= (-1\,|\,\tau)[(-1)^{n}\sec y \cos\{2t(x+y)+\tau z\} \\
&\quad + \sec z \cos(2tx+\tau y)] \\
&\quad + 2(-1)^{n_1}(-1\,|\,\tau_1\tau_2) R_2(x, y, z)\,; \\
F_{3322}(x, y, z) &= (-1\,|\,\tau)[\sec y \cos\{2t(x+y)+\tau z\} \\
&\quad + (-1)^{n}\sec z \cos(2tx+\tau y)] \\
&\quad + 2(-1)^{n_2}(-1\,|\,\tau_1\tau_2) R_2(x, y, z).
\end{aligned}
$$

6. For partition (III) we have the form

$$\phi_{abcd}(x, y, z) = 4\sum_{(m)} q^{m/2}\left(\sum F_{abcd}(x, y, z\,; m)\right).$$

Let

$$
\begin{aligned}
P_3(x, y, z) &= \cos\{(2d_1-t_2)x+(2d_1-\tau_2)y+2\delta_1 z\} \\
&\quad - \cos\{(2d_1+t_2)x+(2d_1+\tau_2)y+2\delta_1 z\}, \\
Q_3(x, y, z) &= \sin\{(2d_1+t_2)x+(2d_1+\tau_2)y+2\delta_1 z\} \\
&\quad + \sin\{(2d_1-t_2)x+(2d_1-\tau_2)y+2\delta_1 z\}.
\end{aligned}
$$

Then

$$
\begin{aligned}
F_{1001}(x, y, z) &= \{\cot(x+y)+\cot z\}\sin(tx+\tau y) \\
&\quad + 2P_3(x, y, z)\,; \\
F_{2301}(x, y, z) &= (-1\,|\,t)\{\cot z-\tan(x+y)\}\cos(tx+\tau y) \\
&\quad + 2(-1)^{d_1}(-1\,|\,t_2) Q_3(x, y, z)\,; \\
F_{1331}(x, y, z) &= (-1\,|\,m)\{\cot(x+y)+\cot z\}\sin(tx+\tau y) \\
&\quad + (-1\,|\,m_2) P_3(x, y, z)\,; \\
F_{2002}(x, y, z) &= \{\cot(x+y)-\tan z\}\sin(tx+\tau y) \\
&\quad + 2(-1)^{\delta_1} P_3(x, y, z)\,; \\
F_{1032}(x, y, z) &= (-1\,|\,\tau)\{\tan(x+y)+\tan z\}\cos(tx+\tau y) \\
&\quad - 2(-1)^{d_1+\delta_1}(-1\,|\,\tau_2) Q_3(x, y, z)\,; \\
F_{2332}(x, y, z) &= (-1\,|\,m)\{\cot(x+y)-\tan z\}\sin(tx+\tau y) \\
&\quad + 2(-1)^{\delta_1}(-1\,|\,m_2) P_3(x, y, z).
\end{aligned}
$$

7. The form in which the partition is (IV) is

$$
\begin{aligned}
\phi_{abcd}(x, y, z) &= T_{abcd}(x, y, z) \\
&\quad + 4\sum_{(m)} q^{m}\left(\sum F_{abcd}(x, y, z\,; m)\right) + 4\sum_{(n)} q^{2n}\left(\sum G_{abcd}(x, y, z\,; n)\right).
\end{aligned}
$$

We have

$$T_{0110}(x, y, z) = T_{3113}(x, y, z) = \csc x \csc y;$$
$$T_{0220}(x, y, z) = T_{3223}(x, y, z) = \sec x \sec y;$$
$$T_{3120}(x, y, z) = T_{0123}(x, y, z) = \csc x \sec y.$$

Let

$$P_4(x, y, z; t_1, \tau_1, d_2, \delta_2) = P_4(x, y, z),$$
$$P_4(x, y, z; t'_1, \tau'_1, d'_2, \delta'_2) = P'_4(x, y, z),$$
$$Q_4(x, y, z; t_1, \tau_1, d_2, \delta_2) = Q_4(x, y, z),$$
$$Q_4(x, y, z; t'_1, \tau'_1, d'_2, \delta'_2) = Q'_4(x, y, z);$$
$$\begin{aligned} P_4(x, y, z) = {} & \cos\{(\tau_1 - 2d_2)x + (\tau_1 - 2\delta_2)y + 2t_1z\} \\ & - \cos\{(\tau_1 + 2d_2)x + (\tau_1 + 2\delta_2)y + 2t_1z\}, \end{aligned}$$
$$\begin{aligned} Q_4(x, y, z) = {} & \sin\{(\tau_1 + 2d_2)x + (\tau_1 + 2\delta_2)y + 2t_1z\} \\ & - \sin\{(\tau_1 - 2d_2)x + (\tau_1 - 2\delta_2)y + 2t_1z\}. \end{aligned}$$

Then

$$\begin{aligned}
F_{0110}(x, y, z) = {} & (\cot x + \cot y)\sin\{2tz + \tau(x + y)\} \\
& + 2P_4(x, y, z); \\
G_{0110}(x, y, z) = {} & (\cot x + \cot y)\sin\{2t'z + \tau'(x + y)\} \\
& + \csc(x + y)\sin 2(dx + \delta y) + 2P'_4(x, y, z); \\
F_{3120}(x, y, z) = {} & (-1|\tau)(\cot x - \tan y)\cos\{2tz + \tau(x + y)\} \\
& + 2(-1)^{\delta_2}(-1|\tau_1)Q_4(x, y, z); \\
G_{3120}(x, y, z) = {} & (-1|\tau')(\cot x - \tan y)\cos\{2t'z + \tau'(x + y)\} \\
& + (-1)^{\delta}\sec(x + y)\sin 2(dx + \delta y) \\
& + 2(-1)^{\delta'_2}(-1|\tau'_1)Q'_4(x, y, z); \\
F_{0220}(x, y, z) = {} & (\tan x + \tan y)\sin\{2tz + \tau(x + y)\} \\
& - 2(-1)^{d_2+\delta_2}P_4(x, y, z); \\
G_{0220}(x, y, z) = {} & (\tan x + \tan y)\sin\{2t'z + \tau'(x + y)\} \\
& - (-1)^{d+\delta}\csc(x + y)\sin 2(dx + \delta y) \\
& - 2(-1)^{d'_2+\delta'_2}P'_4(x, y, z); \\
F_{3113}(x, y, z) = {} & -(\cot x + \cot y)\sin\{2tz + \tau(x + y)\} \\
& - 2P_4(x, y, z); \\
G_{3113}(x, y, z) = {} & (\cot x + \cot y)\sin\{2t'z + \tau'(x + y)\} \\
& + \csc(x + y)\sin 2(dx + \delta y) \\
& + 2P'_4(x, y, z); \\
F_{0123}(x, y, z) = {} & -(-1|\tau)(\cot x - \tan y)\cos\{2tz + \tau(x + y)\} \\
& - 2(-1)^{\delta_2}(-1|\tau_1)Q_4(x, y, z); \\
G_{0123}(x, y, z) = {} & (-1|\tau')(\cot x - \tan y)\cos\{2t'z + \tau'(x + y)\} \\
& + (-1)^{\delta}\sec(x + y)\sin 2(dx + \delta y) \\
& + 2(-1)^{\delta'_2}(-1|\tau'_1)Q'_4(x, y, z);
\end{aligned}$$

$$
\begin{aligned}
F_{3223}(x, y, z) = & -(\tan x + \tan y)\sin\{2tz + \tau(x+y)\} \\
& + 2(-1)^{d_2+\delta_2} P_4(x, y, z);\\
G_{3223}(x, y, z) = & (\tan x + \tan y)\sin\{2t'z + \tau'(x+y)\} \\
& -(-1)^{d+\delta}\csc(x+y)\sin 2(dx + \delta y) \\
& -2(-1)^{d'_2+\delta'_2} P'_4(x, y, z).
\end{aligned}
$$

8. Finally, those expansions in which the partition is (V) have the form

$$
\phi_{abcd}(x, y, z) = 4\sum_{(m)} q^{m/2}\left(\sum F_{abcd}(x, y, z; m)\right).
$$

On putting

$$
\begin{aligned}
P_5(x, y, z) &= A_1(x, y, z), \quad Q_5(x, y, z) = A_2(x, y, z),\\
R_5(x, y, z) &= B_1(x, y, z), \quad S_5(x, y, z) = B_2(x, y, z),\\
A_k(x, y, z) &= \cos\{(t_1 - 2t_2)x + (t_1 - \tau_2)y + \tau_1 z\}\\
&\quad + (-1)^k \cos\{(t_1 + 2t_2)x + (t_1 + \tau_2)y + \tau_1 z\},\\
B_k(x, y, z) &= \sin\{(t_1 + 2t_2)x + (t_1 + \tau_2)y + \tau_1 z\}\\
&\quad + (-1)^k \sin\{(t_1 - 2t_2)x + (t_1 - \tau_2)y + \tau_1 z\},
\end{aligned}
$$

we have

$$
\begin{aligned}
F_{1010}(x, y, z) = & \csc y \sin\{t(x+y) + \tau z\} + 2P_5(x, y, z);\\
F_{2310}(x, y, z) = & (-1\,|\,t)\csc y \cos\{t(x+y) + \tau z\}\\
& + 2(-1)^{n_2}(-1\,|\,t_1) R_5(x, y, z);\\
F_{2020}(x, y, z) = & (-1\,|\,t)\sec y \cos\{t(x+y) + \tau z\}\\
& + 2(-1\,|\,t_1\tau_2) Q_5(x, y, z);\\
F_{1320}(x, y, z) = & \sec y \sin\{t(x+y) + \tau z\}\\
& + 2(-1)^{n_2}(-1\,|\,\tau_2) S_5(x, y, z);\\
F_{2013}(x, y, z) = & (-1\,|\,\tau)\csc y \cos\{t(x+y) + \tau z\}\\
& + 2(-1\,|\,\tau_1) R_5(x, y, z);\\
F_{1313}(x, y, z) = & (-1\,|\,m)\csc y \sin\{t(x+y) + \tau z\}\\
& + 2(-1)^{n_2}(-1\,|\,m_1) P_5(x, y, z);\\
F_{1023}(x, y, z) = & (-1\,|\,m)\sec y \sin\{t(x+y) + \tau z\}\\
& + 2(-1\,|\,m_1)(-1\,|\,\tau_2) S_5(x, y, z);\\
F_{2323}(x, y, z) = & (-1\,|\,\tau)\sec y \cos\{t(x+y) + \tau z\}\\
& + 2(-1)^{n_2}(-1\,|\,\tau_1\tau_2) Q_5(x, y, z).
\end{aligned}
$$

Let

$$
\begin{aligned}
U_5(x, y, z) &= C_1(x, y, z), \quad V_5(x, y, z) = C_2(x, y, z);\\
C_k(x, y, z) &= \cos\{(\tau_2 - t_1)x + (\tau_2 - \tau_1)y + 2t_2 z\}\\
&\quad + (-1)^k \cos\{(\tau_2 + t_1)x + (\tau_2 + \tau_1)y + 2t_2 z\}.
\end{aligned}
$$

Then we have

$$
\begin{aligned}
F_{3030}(x, y, z) &= (-1\,|\,\tau)\sec(x+y)\cos(tx+\tau y) \\
&\quad + 2(-1\,|\,\tau_1\tau_2)V_5(x, y, z)\,; \\
F_{0330}(x, y, z) &= (-1\,|\,m)\csc(x+y)\sin(tx+\tau y) \\
&\quad + 2(-1\,|\,m_1)U_5(x, y, z)\,; \\
F_{3003}(x, y, z) &= \csc(x+y)\sin(tx+\tau y) \\
&\quad + 2(-1)^{n_2}U_5(x, y, z)\,; \\
F_{3333}(x, y, z) &= (-1\,|\,m)\csc(x+y)\sin(tx+\tau y) \\
&\quad + 2(-1)^{n_2}(-1\,|\,m_1)U_5(x, y, z)\,; \\
F_{0000}(x, y, z) &= \csc(x+y)\sin(tx+\tau y) \\
&\quad + 2U_5(x, y, z)\,; \\
F_{0303}(x, y, z) &= (-1\,|\,t)\sec(x+y)\cos(tx+\tau y) \\
&\quad + 2(-1)^{n_2}(-1\,|\,t_1\tau_2)V_5(x, y, z).
\end{aligned}
$$

University of British Columbia,
Vancouver, Canada.

SOME ARITHMETICAL APPLICATIONS OF RESIDUATION.

By Morgan Ward.

1°. The operation of residuation was apparently first considered by Dedekind in his theory of the modules in a ring of algebraic integers [1].[1] It was introduced into polynomial ideal theory by Emanuel Lasker [2], and has since been used systematically by F. S. Macaulay [3] and others. I propose to show here how the operation may be applied to various arithmetical problems,[2] in particular to developing a systematic calculus for the periods of elements in any finite Abelian group.

2°. Consider first for simplicity a cyclic group $\mathfrak{G}$ of order N written additively. Every element α of $\mathfrak{G}$ may be uniquely represented as

$$(2.1) \qquad \alpha = a\gamma, \qquad 0 < a \leqq N$$

where γ is a fixed primitive element of $\mathfrak{G}$ and $a\gamma$ means $\gamma + \gamma + \cdots + \gamma$ taken a times. We write L_α for a in (2.1); for example $L_0 = N$. Let P_α denote the period of α; that is the least positive integer p such that $p\alpha = 0$. The starting point of our investigation is the observation that *P_α is the residual of L_α with respect to N.*

L_ξ considered as an operation on $\mathfrak{G}$ to the finite ring K_N of the integers modulo N is linear and distributive:

$$L_{m\alpha+n\beta} \equiv mL_\alpha + nL_\beta \pmod{N}, \qquad m,\ n \text{ integers.}$$

On the other hand, P_ξ is neither a linear nor a distributive operation; given P_α and P_β, all we can assert about $P_{\alpha+\beta}$ is that it divides $[P_\alpha, P_\beta]$, the least

[1] The numbers [1], [2], $\cdots$ in square brackets refer to the bibliography at the close of the paper.

[2] For example, consider the problem of solving

$$(1) \qquad AX \equiv 0 \ (\text{modd } \mathfrak{m}, F).$$

Here A and F are given polynomials in indeterminates $x_1, \cdots, x_s$ with coefficients in a commutative ring $\mathfrak{R}$ while $\mathfrak{m}$ is an ideal of $\mathfrak{R}$. We seek all solutions X in the quotient ring $\mathfrak{R}[x_1, \cdots, x_s]/\mathfrak{m}$. If $\mathfrak{A}$ and $\mathfrak{F}$ denote the ideals $(\mathfrak{m}, A)$, $(\mathfrak{m}, F)$ then the totality of such solutions of (1) constitute the residual ideal of $\mathfrak{A}$ with respect to $\mathfrak{F}$ (Van der Waerden [4] Chapter XII). Thus the solution of (1) is equivalent to specifying this residual, say by determining a basis for it. I have given a complete solution for the case when $s = 1$ and $\mathfrak{R}$ is the ring of rational integers. Ward [5].

common multiple of P_α and P_β.[3] Simple numerical examples show that $P_{\alpha+\beta}$ may be any divisor whatever of $[P_\alpha, P_\beta]$.

These facts suggest that we introduce in $\mathfrak{G}$ one or more new operations $\xi \circ \eta$, $\xi \times \eta$ such that $P_{\alpha\circ\beta}$, $P_{\alpha\times\beta}$ may be calculated knowing only the values of P_α, P_β; the definition of P_α as a residual immediately suggests how these operations should be defined. But before introducing these operations, we shall briefly summarize the properties of residuation of which we make use.

3. Let $\mathfrak{D}$ be the set of ideals[4] $A, B, C, \cdots$ of a fixed commutative ring containing a unit element. If A and B are any two elements of $\mathfrak{D}$, the residual of B with respect to A is by definition an ideal C such that

$$A \supset BC; \quad \text{if} \quad A \supset BX \quad \text{then} \quad C \supset X.$$

We write as usual $C = A : B$. The residual always exists and has the following properties:

$$\begin{aligned}
(3.1) \qquad & A : B = A : (A, B) = [A, B] : B, \\
& (A : B) : C = (A : C) : B = A : BC, \\
& A = M : N \text{ and } B = M : (M : N) \text{ imply } B = M : A,\ A = M : B, \\
& M : (A_1, A_2, \cdots, A_k) = [M : A_1, M : A_2, \cdots, M : A_k], \\
& [A_1, A_2, \cdots, A_k] : M = [A_1 : M, A_2 : M, \cdots, A_k : M].
\end{aligned}$$

If we restrict $\mathfrak{D}$ to be a principal ideal ring, then $A \supset B$ if and only if there exists a quotient $Q = A/B$ such that $A = QB$. Furthermore this quotient is unique. It is easily shown that $A : B = A/B$ whenever the quotient A/B exists, so that formula (3.1) becomes

$$(3.11) \qquad A : B = \frac{A}{(A, B)} = \frac{B}{[A, B]}.$$

On using this result and the unicity of the quotient, we easily find that the following additional rules for residuation hold in any principal ideal ring.[5]

$$\begin{gathered}
(M, N) = M : (M : N), \qquad M : AB = \{(M : A)(M : B)\} : M, \\
(A_1, A_2, \cdots, A_k) : M = (A_1 : M, \cdots, A_k : M), \\
M : [A_1, A_2, \cdots, A_k] = (M : A_1, \cdots, M : A_k).
\end{gathered}$$

[3] We use $(A, B, \cdots)$, $[A, B, \cdots]$ both for the union and join of ideals $A, B, \cdots$ or the greatest common divisor and least common multiple of integers $A, B, \cdots$.

[4] We use the notation of van der Waerden [4], chapter XII save that roman capitals are used for ideals instead of gothic capitals.

[5] A detailed analysis of the properties of residuation, is given in Ward [6], Dilworth [7].

4°. The formulas of section **3°** give the fundamental relations

$$(4.1)\quad P_a = N : L_a = \frac{N}{(N, L_a)} = \frac{[N, L_a]}{L_a},$$

$$P_a = N : (N : P_a),\ L_a = N : P_a \text{ if and only if } L_a \text{ divides } N.$$

We define our new operations over the group $\mathfrak{G}$ as follows. We write

$$\delta = (\xi, \eta) \text{ if } L_\delta = (L_\xi, L_\eta),$$
$$\mu = [\xi, \eta] \text{ if } L_\mu \equiv [L_\xi, L_\eta] \pmod{N}.$$

It is clear that the group $\mathfrak{G}$ forms an arithmetic structure [6] with respect to the operations of union and cross-cut thus defined which is simply isomorphic with the structure of the ring K_N.

The third operation over $\mathfrak{G}$ which we shall consider is a multiplication simply isomorphic with multiplication in K_N: we write

$$\pi = \xi \cdot \eta \text{ if } L_\pi \equiv L_\xi L_\eta \pmod{N}.$$

If we call two elements of $\mathfrak{G}$ equivalent if and only if each divides the other, then equivalent elements have the same period and conversely.

The periods of δ, μ and π obey the following simple rules which are easy consequences of (4.1) and the formulas of section **3°**:

$$P_{(\xi,\eta)} = [P_\xi, P_\eta], \quad P_{(\xi_1, \xi_2, \ldots, \xi_k)} = [P_{\xi_1}, P_{\xi_2}, \cdots, P_{\xi_k}],$$
$$P_{[\xi,\eta]} = (P_\xi, P_\eta), \quad P_{[\xi_1, \xi_2, \ldots, \xi_k]} = (P_{\xi_1}, P_{\xi_2}, \cdots, P_{\xi_k}),$$
$$P_{\xi \cdot \eta} = \{P_\xi P_\eta\} : N, \ P_{\xi_1 \cdot \xi_2 \cdot \ldots \cdot \xi_k} = \{P_{\xi_1} P_{\xi_2} \cdots P_{\xi_k}\} : N^{k-1}.$$

Thus for each operation the period is readily calculated from the period of its constituents. It is possible to define a residual $\xi : \eta$ in $\mathfrak{G}$ by $L_{\xi:\eta} = L_\xi : L_\eta$, but its period is not calculable in terms of the periods of ξ and η alone.

5°. If we choose a different primitive element γ' in place of γ defining a new operator L'_ξ over $\mathfrak{G}$ we have

$$L'_a \equiv L'_\gamma L_a \pmod{N}, \qquad L_a \equiv L_{\gamma'} L'_a \pmod{N}$$

where $L'_\gamma L_{\gamma'} \equiv 1 \pmod{N}$, so that both L'_γ and $L_{\gamma'}$ are prime to N. It readily follows that the operations (ξ, η), $[\xi, \eta]$ are independent of the particular base γ chosen to define them. The situation for the product is different. We

[6] Or distributive lattice. See Ore [8] or Ward [6] for detailed definition.

find that $(\alpha\cdot\beta)'=\gamma'\cdot(\alpha\cdot\beta)$. On the other hand $P'_\alpha=P_\alpha$. The formulas (4.1) are thus unchanged. For example:

$$P'_{(\xi\,.\,\eta)'}=P_{\gamma'\,.\,(\xi\,.\,\eta)}=\{P_{\gamma'}P_{\xi\,.\,\eta}\}:N=\{N\,.\,P_{\xi\eta}\}:N=P_{\xi\eta}.$$

6°. Suppose now that the group $\mathfrak{G}$ is the direct sum of K cyclic groups $\mathfrak{G}^{(1)},\cdots,\mathfrak{G}^{(k)}$ of orders $N^{(1)},\cdots,N^{(k)}$:

$$\mathfrak{G}=\mathfrak{G}^{(1)}\dotplus\mathfrak{G}^{(2)}\dotplus\cdots\dotplus\mathfrak{G}^{(i)}\dotplus\cdots\dotplus\mathfrak{G}^{(k)}, \tag{6.1}$$

so that the typical element α of $\mathfrak{G}$ is of the form

$$\alpha=\alpha^{(1)}+\alpha^{(2)}+\cdots+\alpha^{(i)}+\cdots+\alpha^{(k)}.$$

We select in each group $\mathfrak{G}^{(i)}$ a primitive element $\gamma^{(i)}$ and define operators $L_\alpha{}^{(i)}$ as in section **2°** by

$$L_\alpha{}^{(i)}=a^{(i)},\qquad \alpha^{(i)}=a^{(i)}\gamma^{(i)},\qquad\qquad (i=1,2,\cdots,k).$$

We then associate with the element α the vector $\mathfrak{L}_\alpha$ whose i-th component is $L_\alpha{}^{(i)}$. The operations (α,β), $[\alpha,\beta]$ $\alpha\cdot\beta$ over $\mathfrak{G}$ of union, cross-cut and product are defined by the vectors $\mathfrak{L}_{(\alpha,\beta)}$, $\mathfrak{L}_{[\alpha,\beta]}$, $\mathfrak{L}_{\alpha\,.\,\beta}$ with components $(L_\alpha{}^{(i)},L_\beta{}^{(i)})$, $[L_\alpha{}^{(i)},L_\beta{}^{(i)}]$, $L_\alpha{}^{(i)}L_\beta{}^{(i)}$ respectively where the components are taken modulo $N^{(i)}$ in the associated rings $K_{N^{(i)}}$.

With an obvious extension of notation, we write

$$\mathfrak{L}_{(\alpha,\beta)}=(\mathfrak{L}_\alpha,\mathfrak{L}_\beta),\qquad \mathfrak{L}_{[\alpha,\beta]}=[\mathfrak{L}_\alpha,\mathfrak{L}_\beta]$$
$$\mathfrak{L}_{\alpha\,.\,\beta}=\mathfrak{L}_\alpha\cdot\mathfrak{L}_\beta.$$

The *vectorial period* of α is defined as the vector $\mathfrak{P}_\alpha$ with components $P_\alpha{}^{(i)}=N^{(i)}:L^{(i)}$. If $\mathfrak{N}$ denotes the vector with components $N^{(1)},N^{(2)},\cdots,N^{(k)}$ ($\mathfrak{N}$ is simply $\mathfrak{L}_0$, where 0 is the identity element of $\mathfrak{G}$) then we write

$$\mathfrak{P}_\alpha=\mathfrak{N}:\mathfrak{L}_\alpha.$$

These definitions allow us to extend immediately the formulas of section **4°**; thus

$$\mathfrak{P}_{[\alpha,\beta]}=(\mathfrak{P}_\alpha,\mathfrak{P}_\beta),\qquad \mathfrak{P}_{(\alpha,\beta)}=[\mathfrak{P}_\alpha,\mathfrak{P}_\beta],\qquad \mathfrak{P}_{\alpha\,.\,\beta}=\mathfrak{P}_\alpha\cdot\mathfrak{P}_\beta:\mathfrak{N}.$$

The actual period of α (that is, the least positive integer p such that $p\alpha=0$) is simply the least common multiple of the components of the vectorial period. Denoting it as before by P_α, we have

$$P_\alpha=[P_\alpha{}^{(1)},P_\alpha{}^{(2)},\cdots,P_\alpha{}^{(k)}].$$

We cannot calculate the scalar period of $\alpha \cdot \beta$ or $[\alpha, \beta]$ directly in terms of the scalar periods of α and β. But for the union (α, β) we have the elegant formula

$$P_{(\alpha,\beta)} = [P_\alpha, P_\beta].$$

These considerations apply to any finite Abelian group since every such group may be represented as a direct sum of cyclic groups. If we assume as is always possible that the order of each summand is a power of a prime, then the number of summands $\mathfrak{G}^{(i)}$ is uniquely specified and also the orders $^{N(i)}$.

To remove in part the ambiguity in the definition of the components of $\mathfrak{L}_\xi$ and $\mathfrak{P}_\xi$ due to the fact that the order of the groups $\mathfrak{G}^{(i)}$ in (6.1) is unspecified, we agree to arrange the prime power orders $N^{(i)}$ first in the natural order of the primes, and then arrange the powers of each prime in order of magnitude. The remaining ambiguity in the order of the components due to adjoining isomorphic groups in the decomposition (6.1) appears to be inherent, as the set of all vector functions $\mathfrak{P}_\xi$ over $\mathfrak{G}$ can be regarded as a basis for a representation of the group of automorphisms of $\mathfrak{G}$, each function being corollated with the sub-group of automorphisms leaving its components unchanged in order, but changing possibly the basis elements $\gamma^{(i)}$ in terms of which the components $L_\xi^{(i)}$ are specified. We have already seen in section **5**° that the components of $\mathfrak{P}_\xi$ are unaffected by such changes of base. The remaining automorphisms of $\mathfrak{G}$ will permute isomorphic groups in (6.1) thus inducing a permutation of the vector functions $\mathfrak{P}_\xi$. In any event the scalar period function P_ξ remains unaffected.

Since any finite field excluding its zero element is a cyclic group with respect to multiplication, the calculus we have developed in section **4**° carries over entire to the periods of elements in any such field. The vectorial calculus of the present section similarly applies to the periods of the units in any finite commutative ring.

7°. The operations which we have defined over the finite group have analogues in common arithmetic. For if

$$A = \prod_1^\infty P_n^{a_n}, \qquad B = \prod_1^\infty P_n^{b_n}$$

are the decompositions of the positive integers A and B into their prime factors, where $P_1, P_2, P_3, \cdots$ denote the primes $2, 3, 5, \cdots$ in their natural order and only a finite number of the exponents a_n, b_n are not zero, we may define a "union," "cross-cut" and "product" of A and B "of the second kind" by

$$(A, B) = \prod_1^\infty P_n{}^{(a_n, b_n)} \qquad [A, B] = \prod_1^\infty P_n{}^{[a_n, b_n]}$$

$$A \cdot B = \prod_1^\infty P_n{}^{a_n, b_n}.$$

The product of the second kind is distributive with respect to the ordinary product $A \times B$ or "product of the first kind":

$$A \cdot (B \times C) = (A \cdot B) \times (A \cdot C).$$

The analogy with our treatment of groups becomes evident if we think of A as specified by the vector with components $a_1, a_2, \cdots$.

Indeed our product is the arithmetical analogue of the "multiplication of the second order" considered by De Morgan [9] and others [10] in the hierarchy of operations

$$A + B,\ A \times B = \exp(\log A + \log B),\ A \cdot B = \exp(e^{\log\log A + \log\log B}) = A^{\log B}, \cdots.$$

PASADENA, CALIFORNIA.

REFERENCES.

1. R. Dedekind, Supplement XI of Dirichlet's *Vorlesungen*, 4th edition (1894), § 170; *Collected Works*, vol. III, p. 71.
2. E. Lasker, *Mathematische Annalen*, vol. 60 (1905), pp. 20-116.
3. F. S. Macaulay, *The Algebraic Theory of Modular Systems*, Cambridge (1916).
4. B. L. van der Waerden, *Moderne Algebra*, vol. 2 (Berlin, 1931).
5. M. Ward, *Transactions of the American Mathematical Society*, vol. 35 (1933), pp. 254-260.
6. M. Ward, Paper submitted to the *Duke Journal*.
7. R. P. Dilworth, Paper submitted to the *Transactions of the American Mathematical Society*.
8. O. Ore, *Annals of Mathematics*, vol. 36 (1935), pp. 406-437.
9. A. De Morgan, *Trigonometry and Double Algebra*, p. 166.
10. R. E. Moritz, *Tohôku Journal*, vol. 21 (1921), pp. 51-64.

RATIONAL CURVES OF ORDER $n + 2$ INVARIANT UNDER DIHEDRAL COLLINEATION GROUPS OF ORDER $2n$.*

By R. M. Winger.

1. Introduction. The problem of self-projective curves, i. e. curves which admit linear transformations into themselves, may be said to go back to Euler and Steiner who respectively investigated curves with a diameter and a center. Projectively considered, it is a matter of indifference whether one asks that a curve have a center or a diameter. For a center of symmetry is merely the center of a harmonic homology, or reflexion, whose axis is the line at infinity; while a diameter is an axis of reflexion whose center is a point at infinity. The diameter becomes an axis of symmetry when the center of the reflexion is in the direction of the perpendicular. An algebraic curve may have several axes of symmetry although it can have but a single center. If there are two axes, the product of the corresponding reflexions will be a rotation of finite period, say p, and the reflexions will generate a dihedral group of order $2p$ which leaves the curve invariant.[1]

Multiple axial symmetry of an algebraic curve is thus intimately connected with dihedral groups. Indeed the whole question of the symmetry of curves or other geometrical figures is but a metrical aspect of collineation groups.[2]

The problem of self-projective curves, consciously formulated, has been attacked by numerous writers, including Klein and Lie, S. Kantor, Wiman, Ciani, Snyder and the Author.

The maximum axial symmetry of an algebraic curve of order m is m-fold, the axes radiating from a point at equal angles.[3] A rational m-ic with $m - 1$ axes of symmetry must be of even order and a trochoid. There is in fact a one-parameter family of such curves. There is also a one-parameter family of trochoids possessing $m - 2$ axes of symmetry when m is even. These two

* Read before the American Mathematical Society, September 3, 1936. Received by the Editors May 24, 1937.

[1] Winger, *American Mathematical Monthly*, vol. 37 (1930), p. 5. Also Loria, *Spezielle Algebraische und Transzendente Ebene Kurven*, Second Edition, vol. I, Chapter 10.

[2] Cf. Jaeger, *Lectures on the Principle of Symmetry*, Amsterdam, 1920, Chapter 3.

[3] Winger, *American Mathematical Monthly*, *loc. cit.*, also *American Journal of Mathematics*, vol. 36 (1914), p. 66 where the equations are obtained for rational curves.

systems have been discussed elsewhere.[4] For all values of m $(m > 3)$, even or odd however there is a one-parameter family of non-trochoidal rational curves with $m - 2$ axes of symmetry. These form the subject of the present study. We shall find it convenient to take $m = n + 2$. Projectively stated then, the problem is the consideration of those rational curves of order $n + 2$ (other than projective trochoids) which are invariant under dihedral collineation groups of order $2n$. The curves divide broadly into two classes, according as n is odd or even, corresponding to the two main species of dihedral groups. The family includes as special cases many remarkable individuals among which may be mentioned the self-dual curves of Wear,[5] which are autopolar with respect to the maximum number of conics; certain of the polar tangent curves considered by Stratton.[6] A metric version of Wear's curves are the self-dual rational curves of maximum symmetry studied by Duncan (for n odd).[7]

As usual there are two groups involved, the binary group on the parameter and the ternary group on the points of the curve. The binary group is generated by the collineation

$$s: \quad t' = \epsilon t, \qquad \epsilon^n = 1,$$

and the involution

$$r: \quad t' = 1/t.$$

The first generates an invariant cyclic subgroup g_n whose elements are $t' = \epsilon^i t$, $i = 1, 2, \cdots n$. The product of these by r yields the n involutions $t' = \epsilon^i / t$.

The ternary group is generated by the cyclic substitution

$$S: \quad x'_1 = \epsilon^2 x_1, \ x'_2 = x_2, \ x'_3 = \epsilon x_3$$

or

$$x'_1 = \epsilon x_1, \ x'_2 = \epsilon^{-1} x_2, \ x'_3 = x_3$$

of period n, which generates a ternary cyclic subgroup G_n, and the reflexion

$$R: \quad x'_1 = x_2, \ x'_2 = x_1, \ x'_3 = x_3.$$

The elements of the ternary group comprise the elements of the cyclic

$$G_n: \quad x'_1 = \epsilon^i x_1, \ x'_2 = \epsilon^{-i} x_2, \ x'_3 = x_3, \qquad (i = 1, 2, \cdots n),$$

[4] *American Mathematical Monthly*, vol. 39 (1932), p. 578.

[5] *American Journal of Mathematics*, vol. 51 (1929), p. 482.

[6] *American Mathematical Monthly*, vol. 43 (1936), p. 398.

[7] *Bulletin of the American Mathematical Society*, vol. 40 (1934), p. 344.

together with the products of these and R which are the n reflexions:

$$x'_1 = \epsilon^i x_2,\ x'_2 = \epsilon^{-i} x_1,\ x'_3 = x_3.$$

The axes of these reflexions are $x_1 - \epsilon^i x_2 = 0$, while their corresponding centers are $(\epsilon^i, -1, 0)$. The centers thus lie on the line $x_3 = 0$ and the axes meet at $u_3 = 0$. The equations of the axes as a whole are $x_1^n - x_2^n = 0$.

There is a fundamental difference according as n is odd or even. When n is odd the axes are all conjugate (equivalent), and no center lies on an axis. But when n is even, say $n = 2k$, they divide into two complementary sets $A+$, $A-$:

$$A+:\quad x_1^k + x_2^k = 0, \qquad A-:\quad x_1^k - x_2^k = 0,$$

the corresponding centers of which we shall denote respectively by $C+$ and $C-$. Now the axes (and centers) of either set are conjugate among themselves only. Further the axes and centers are incident in pairs: Each center $C\pm$ lies on an axis $A\mp$ of the complementary set when k is odd but on an axis $A\pm$ of its own set when k is even. Again when n is even the binary group g_n contains an extra involution, $t' = -t$, and the ternary G_n an extra reflexion with center u_3 and axis x_3. The special sets of conjugate points of the binary group are the pair of parameters 0, ∞ and the two sets of n,

$$t^n + 1 = 0,\ t^n - 1 = 0.$$

I. The General Curve of the Pencil.

2. Consider now a rational curve of order $n+2$,

$$x_i = f_i(t), \qquad (i = 1, 2, 3),$$

where f_i are polynomials of order $n+2$, and suppose that when t is transformed into t', x (parameter t) is carried into x' (parameter t'). Then apply to f_i the generating transformations of the binary group and ask that x_i be transformed according to the generators of the ternary group. We find thus the canonical equations of the most general rational curve of order $n+2$ invariant under our group:[8]

$$x_1 = t^{n+2} + at^2,\ x_2 = at^n + 1,\ x_3 = t^{n+1} + t, \tag{1}$$

where a is an arbitrary constant. The line equations are

[8] When n is even and $n/2+1$ odd, there is a second variety in which $x_3 = t^{n/2+1}$. This is the trochoidal case already discussed, *American Mathematical Monthly*, vol. 39 (1932), p. 584.

$$
(2) \qquad \begin{aligned} u_1 &= at^{2n} + [n+1-(n-1)a]t^n + 1 \\ u_2 &= t^{2n+2} + [n+1-(n-1)a]t^{n+2} + at^2 \\ u_3 &= -2at^{2n+1} + [a^2(n-2)-n-2]t^{n+1} - 2at. \end{aligned}
$$

We shall first consider the properties of the curves for general parameter a, distinguishing the two cases n odd, even. While the curves are highly restricted, the treatment is quite general. We deal in a sense with a doubly infinite system of curves, since there is a one-parameter family for each value of n, which in turn may be any positive integer.[9]

The flex form is seen to factor into

$$
(3) \qquad (t^n+1)\{2at^{2n} + [(n-1)(n-2)a^2 - 2n^2a + (n+1)(n+2)]t^n + 2a\}.
$$

An immediate observation is that the first factor represents one of the special sets of conjugate points (parameters) of the binary group and it is obvious from (1) that they are cut out by x_3, hence

One third of the flexes, whose parameters comprise one of the special sets of conjugate points of the binary group, lie on the fixed line of the ternary group. The other intersections 0, ∞ of x_3 constitute the special set of two conjugate points. The remaining flexes form a general conjugate set under the binary group, since obviously the other special set t^n-1 is not a factor of the flex form.

3. The general curve, n odd. The double points of the involution $t' = 1/t$ are ± 1. By referring to equations (1) we see that the parameter 1 is cut out by the axis $x_1 - x_2 = 0$. The parameter -1 names a flex whose coördinates are $(1, -1, 0)$, a center of reflexion. Further the tangent at the point $t=1$ is, from (2), $x_1 + x_2 - (a+1)x_3 = 0$ and thus goes through the center. Since the centers as well as the axes are conjugate we may say

The centers of reflexion are flexes and thus lie on the curve. From each runs one simple tangent to the curve whose contact is on the corresponding axis. These contacts as a whole, namely $t^n - 1 = 0$ comprise the other special set of parameters under the binary group.

We have noted one point on each axis. The others must be fixed under a reflexion, while their parameters are paired under an involution. This

[9] When $n=1$ we get a family of cubics which are all projectively equivalent except the cuspidal curve, $a = -3$.

requires nodes or multiple points of higher order. Anything but nodes however would demand more double points than are available. Hence

On each axis lie $(n+1)/2$ nodes, which accounts for all of them.

The nodes fall into $(n+1)/2$ special conjugate sets under the ternary group, the parameters of each set however forming general conjugate sets of the binary group. We are now in possession of all special sets of conjugate points of the ternary group which lie on the curve: Besides the nodes they are the three sets whose parameters are the special sets of the binary group.

Again, all tangents from a center of reflexion are fixed under the reflexion but their contacts are subject to the corresponding involution. This means in the general case that all tangents from a center, except the flex tangent of the center itself and the simple tangent already noted, must be bitangents. We may say then, the curve being of class $2n+2$,

$n-1$ bitangents meet at each center of reflexion, accounting for half the total.

4. The general curve, *n* even. When n is even the changes we have to record are due mainly to the division of the axes (and centers) into two sets which are not conjugate to each other, and to the presence of an additional reflexion, whose center is u_3 and whose axis is x_3. The double points of the associated involution, $t'=-t$, are $0, \infty$, which lie on the axis and name contacts of tangents from the center. The other intersections of this axis will be nodes but, since they are also flexes, biflecnodes are formed. Hence

There are $n/2$ biflecnodes on the axis $x_3=0$.

Now setting $n=2k$, we have to distinguish two cases according as k is odd or even.

k odd: The axis $x_1-x_2=0$ cuts out the pairs of parameters $t^2-1=0$ and $t^2+1=0$, which are double points of two involutions. The first pair name contacts of tangents from the center of the reflexion, while the others are parameters of a biflecnode on x_3. The remaining intersections of the axis are ordinary nodes. The biflecnode is a center of reflexion whose axis is $x_1+x_2=0$. All intersections of this axis are nodes, for the double points of the associated involution fall at the center. And all the tangents from a biflecnode, except the nodal tangents themselves, are double lines. Since these results are typical, we have

Each of the axes $A-$ cuts out one biflecnode, $(n-2)/2$ ordinary nodes, and two contacts of simple tangents from its associated center; whereas each axis $A+$ cuts out $(n+2)/2$ ordinary double points. From each biflecnode run $n-2$ double tangents, but from each of the other centers, including u_3, run two simple and n double tangents.

This accounts for all of the double points and n^2 double lines, leaving $n(n-2)$ double lines to be distributed in general sets.

k even: Each of the axes $A-$ cuts out $n/2$ ordinary double points and two contacts of tangents from the corresponding center, while each of the complementary sets of axes $A+$ cuts out one biflecnode and $n/2$ ordinary double points, accounting for all of the double points.

Each biflecnode is a center $C+$ and thus lies on an axis conjugate to its own. Through each biflecnode pass $n-2$ bitangents and through each of the other centers, including u_3, pass two simple and n double tangents, accounting for a total of n^2 bitangents as before.

5. The pencil of invariant conics. Whether n is odd or even, there is a pencil of conics which are individually invariant under the group. The equation of the pencil in points and in lines may be written

$$x_1x_2=\lambda x_3^2 \text{ and } 4\lambda u_1u_2=u_3^2. \tag{4}$$

All proper members of the pencil touch the curve at the points $t=0, \infty$ and cut out besides $2n$ points which comprise conjugate sets, special or general, under both groups. Likewise the $4n$ common lines of each conic and the curve form conjugate sets of both groups. Indeed the common points and lines of the curve and degenerate members of both pencils belong in conjugate sets of the groups. Conversely, all conjugate sets of points and lines of both groups, special as well as general, are common points and lines of the conics and the curve. Among the conics of particular interest are:

One conic	Value of λ
(5) with contacts at $t^n=1$,	$\dfrac{(a+1)^2}{4}$
(6) on $2n$ flexes,	$\dfrac{na[(n-3)a-(n+3)]}{(n-1)(n-2)a-(n+1)(n+2)}$
(7) on flex lines[10] $t^n+1=0$,	$\dfrac{[(n-2)a+n+2]^2}{4n^2}$

[10] Biflecnodal tangents when n is even.

(8) on the other $2n$ flex lines, $$\frac{na[(n-2)a-(n+2)]}{(n-1)^2a-(n+1)^2}$$

(9) on contacts of tangents[11] from u_3, $$\frac{na(a+1)}{(n-2)a+n+2}.$$

We note that the value of λ in (5) is independent of n. We observe also that two or more of the conics may coincide or one or more of the conics may degenerate ($\lambda=0, \infty$) for particular values of a, i.e. for certain special curves. On the other hand, the coincidence of any two of the conics or the degeneration of any one implies a specialization of the curve.

II. Special Cases.

6. By assigning particular values to a we obtain an infinite system of projectively distinct curves. Among the most interesting special curves are those with singularities arising from the coincidence of the flexes. These are found by equating to zero the discriminant of the second factor of the flex form (3).[12] We shall consider them briefly. The discriminant of the second factor of the flex form, considered as a quadratic, is

$$(a-1)[(n-1)a-(n+1)][(n-2)a-(n+2)] \\ \times [(n-1)(n-2)a-(n+1)(n+2)].$$

The complete discriminant is a power of this, together with two other factors corresponding to $a=0, \infty$. For obviously these values imply multiple roots of the flex equation. When $a=1, \infty$ the curve degenerates for then the three line sections $x_i=0$ have a common factor. Ordinarily the flexes in question form a general set under the binary group. But when the discriminant vanishes the flex parameters coincide at least in pairs and hence must reduce to special sets. We find readily the following consequences of a vanishing discriminant:

When	The curve has
$a=(n+1)/(n-1)$,	undulations at $t^n-1=0$
$a=(n+2)/(n-2)$,	cusps at $t^n-1=0$
$a=(n+1)(n+2)/(n-1)(n-2)$,	5-point contact flexes at $t^n+1=0$
$a=0$,	n-point contact tangents at $t=0, \infty$.

In the first three cases the flexes involved coincide in pairs at the respective

[11] Bitangents when n is even.

[12] The flexes represented by the first factor obviously cannot coincide.

singularities, but when $a = 0$, they coincide n at a time. We shall now discuss the individual curves.

$$a = (n+1)/(n-1).$$

7. The undulations lie on the axes of reflexion, one on each when n is odd and two on each axis $A-$ when n is even. The undulation tangents pass through the centers of reflexion, one through each when n is odd and two through each $C-$ center when n is even. The other properties of the curve noted for the general case must be modified to conform to the fact that an undulation arises from the coincidence of three lines through a center of reflexion, one bitangent, one line joining two flexes, and one simple tangent.

$$a = (n+2)/(n-2).$$

8. This is the self-dual case treated by Wear (l. c.). The cusps reduce the class of the curve to $n+2$ and the number of double lines to $n(n-1)/2$. When n is odd, there is one cusp on each axis, which is the cusp tangent. Also $(n-1)/2$ double lines meet at each flex (center of reflexion). When n is even however the cusps fall in pairs on the axes $A-$, which are now double-cusp tangents and count as double lines. Further when n is even and k is odd, each double-cusp tangent cuts out one of the biflecnodes and $(n-6)/2$ ordinary nodes; while the other axes $A+$ cut out $(n+2)/2$ ordinary nodes each. Further $(n-4)/2$ double lines, including one double cusp tangent, meet at each biflecnode and $(n+2)/2$ double lines meet at each of the other centers $C-$. When k is even, each double cusp tangent cuts out $(n-4)/2$ ordinary nodes, whereas the other axes $A+$ cut out $n/2$ ordinary nodes and one biflecnode each. Hence in this case $(n+2)/2$ double lines, including one double cusp tangent, meet at each center $C-$, while $(n-4)/2$ double lines meet at each center $C+$, which is now a biflecnode. When n is even, the tangents from the additional center u_3 consist of two simple tangents and the k double cusp tangents, which have already been enumerated. All the singularities of the self-dual curves have now been accounted for.

The line equations of the curve, after factoring out the cusp form, reduce to

$$\begin{aligned} u_1 &= (n+2)t^n - (n-2) \\ u_2 &= (n-2)t^{n+2} - (n+2)t^2 \qquad (10) \\ u_3 &= -2(n+2)t(t^n-1). \end{aligned}$$

The binary transformation $t' = \eta/t$, where $\eta^n = -1$ induces the polarity

$$\text{II}: \quad u_1 = x_1, \; u_2 = x_2, \; u_3 = -2a\eta x_3, \qquad a = (n+2)/(n-2)$$

which interchanges the dual singularities of the curve. This combined with the G_{2n} yields the $2n$ correlations:

$$(11)\qquad \begin{aligned} u_1 &= \epsilon^i x_1,\ u_2 = \epsilon^{1-i} x_2,\ u_3 = -2a\eta x_3, \\ u_1 &= \epsilon^i x_2,\ u_2 = \epsilon^{1-i} x_1,\ u_3 = -2a\eta x_3, \end{aligned} \qquad (i = 1, 2, \cdots n).$$

The first set of these are polarities in all cases. Moreover when n is odd, say $n = 2p - 1$, we get a polarity in the second set, namely when $i = p$. We may summarize thus:

When $a = (n+2)/(n-2)$, the curve admits a G_{4n}, comprising $2n$ collineations and $2n$ correlations. Of the correlations n are polarities when n is even while $n+1$ are polarities when n is odd. The curve is thus autopolar with respect to n or $n+1$ conics, according as n is even or odd—the maximum number for a rational curve of order $n+2$.

$$a = (n+1)(n+2)/(n-1)(n-2).$$

9. All of the flexes of this curve unite to form higher flexes with 5-point contact tangents at $t^n + 1 = 0$, distinct when n is odd but joined at special biflecnodes when n is even. These singularities absorb a total of $3n$ double lines and reduce the number of tangents that can be drawn from each of them by 5 or 10, according as n is odd or even. The properties of the general curve will be modified in consequence. Hence, when n is odd the number of double lines that meet at each flex in question is $n-2$, leaving $n(n-3)$, which belong in general sets. When n is even, k odd or even, $n-4$ double lines meet at each biflecnode and n at each of the other $k+1$ centers, $n(n-4)$ remaining.

$$a = 0.$$

10. The equations reduce to

$$(12)\qquad x_1 = t^{n+2},\ x_2 = 1,\ x_3 = t^{n+1} + t.$$

There are hyperosculation points at $t = 0, \infty$, whose tangents are $x_1 x_2 = 0$ and which absorb together $2n$ flexes and $n(n-1)$, i. e. half of the double lines. The other intersections of x_3 are simple flexes or ordinary biflecnodes according as n is odd or even. Thus all the flexes lie on x_3. All tangents from u_3 coincide with $x_1 x_2 = 0$. The pair of lines $x_1 \pm x_2 = 0$ are cyclic cutting out the points $t^{n+2} \pm 1 = 0$.

The dual of the line curve may be written, see equations (2),

$$
(13) \qquad \begin{aligned} x_1 &= t^{2n+2} + (n+1)t^{n+2}, \\ x_2 &= (n+1)t^n + 1, \\ x_3 &= -(n+2)t^{n+1}. \end{aligned}
$$

This is a projective trochoid having an $(n+1)$-fold point with coincident parameters at each of the points $t = 0, \infty$. Each of these points, which is the dual of a tangent with $(n+2)$-point contact, is thus equivalent to n cusps and $n(n-1)/2$ nodes.

Numerous other noteworthy curves occur in the family, among the most interesting of which are those arising when some of the invariant conics of § 5 degenerate or coincide. We get degenerate conics (and special curves) when $\lambda = 0$ or ∞. These cases we shall now consider.

$$a = -1.$$

11. This curve has an n-fold point at u_3 with parameters $t^n - 1 = 0$. When n is odd the multiple point has distinct tangents and absorbs $n(n-1)/2$ double points, leaving n, one of which lies on each axis. Moreover the tangent at each branch is on one of the centers of reflexion. When n is even however, the tangents touch in pairs, forming k tac-nodes within the multiple point which absorb $n/2$ extra double points and a like number of double lines. These with the usual $n/2$ biflecnodes account for all of the nodes. The tac-nodal tangents are the axes $A+$ or $A-$ according as k is odd or even and are thus on the centers $C-$ in both cases. Each counts as a bitangent and replaces the two simple tangents from a center $C-$ of the general case. Hence, k odd or even, all of the tangents from each center $C-$ are double lines. From each center $C+$ (a biflecnode) run $n-2$ double lines. We have now accounted for n^2 bitangents, including those absorbed by the tac-nodes. The center u_3 yields no new ones since all of the double lines from there are the tac-nodal tangents. Conics (5) and (9) in lines both degenerate to $u_3{}^2 = 0$, i. e. the multiple point repeated.

$$a = (n+3)/(n-3).$$

12. Conic (6) reduces to $x_3{}^2 = 0$ when

$$a = (n+1)(n+2)/(n-1)(n-2)$$

for then all the flexes combine to form higher flexes (supra). This conic also degenerates when $a = 0$ or $(n+3)/(n-3)$, reducing to $x_1 x_2 = 0$. The case $a = 0$ has already been noticed. For the other value of a the flex form factors into

$$[(n-3)t^n+n+3][(n+3)t^n+n-3]$$

and we have at once:

If $n>3$ and $a=(n+3)/(n-3)$, the $3n$ flexes are cut out by the sides of the invariant triangle, one third lying on each side. When n is odd all the flexes are distinct but when n is even those on $x_3=0$ form biflecnodes as usual.[13]

$$a=-(n+2)/(n-2).$$

13. Conic (7) in lines degenerates to $u_3{}^2=0$ for this value of a which is the negative of that for the self-dual case. The tangents from u_3 by (2)) are now given by $t(t^n+1)^2=0$, hence in this case

The tangents from the fixed point of the ternary group, except the simple tangents $0, \infty$, are all flex tangents with their contacts on the fixed line. If n is odd they are all simple flex lines, but if n is even each is a tangent at a special biflecnode whose tangents coincide.

Each such biflecnode or oscnode counts for three ordinary double points, i. e. contains two latent bitangents. Since an ordinary double flex tangent is equivalent to two flex and four double tangents, these oscnodal tangents are equivalent to two flex and six double tangents. Thus the tangents from u_3 account for $3n$ bitangents. The oscnodes are centers of reflexion $C+$, lying on the axes $A-$ or $A+$ according as k is odd or even. From each can be drawn to the curve, besides its own tangent, $2n-4$ tangents which comprise $n-2$ double lines; while from each of the centers $C-$ can be drawn 2 simple and n double tangents. We have now accounted for $n(n+2)$ bitangents, $n(n-4)$ remaining.

$$a=(n+1)^2/(n-1)^2.$$

14. For the self-dual curves, as well as when $a=0$, the line conic (8) reduces to $u_3{}^2=0$, i. e. the point (repeated) in which the cusp or the hyperosculation tangents respectively meet. When a takes the value above the conic becomes $u_1u_2=0$, whose common lines with the curve, except $0, \infty$, are given by

$$[(n+1)t^n-n+1]^2[(n-1)t^n-n-1]^2\equiv(\text{second factor of flex form})^2.$$

[13] It is a rather common experience to encounter a dihedral curve with numerous flexes on a line at the centers of reflexion. The trinomial curves $x^n+y^n+z^n=0$, n a positive integer, meet the sides of the reference triangle, which are axes of homology, in hyperosculation points which absorb all of the flexes. But this is the first non-trivial instance I know of a rational curve whose full complement of simple flexes lie on the sides of a proper triangle.

Hence *n flex tangents meet at each of the points* $u_1 = 0$, $u_2 = 0$. Or,

When $a = (n+1)^2/(n-1)^2$, *one third of the inflexional lines meet at each of the points cut out by the line on which lie the contacts of the remaining third of inflexional lines.*

$$a = -(n+1)/(n-1).$$

15. *If n is odd and* $a = -(n+1)/(n-1)$, *the tangents from the centers of reflexion (flexes) touch the curve again on the axes, at the points* $t^n - 1 = 0$.

This fact, first observed for the quintic [14] is easily verified for the general case. These line singularities are triple lines, each equivalent to one flex tangent and two bitangents, accounting thus for a total of $2n$ bitangents. From each center can be drawn $2n-2$ other tangents which must unite to form $n-1$ bitangents. This accounts for $n(n+1)$, so that $n(n-3)$ are left.

What is the meaning when n is even? The biflecnodal tangents cannot touch the curve at one of the points $t^n - 1 = 0$ for these points do not lie on the axes of which the biflecnodes are centers. Now we saw that there is a conic (5) which touches the curve at the points $t^n - 1 = 0$; and when n is odd the flex lines $t^n + 1 = 0$ touch the curve at the same points. Hence the conic (7) on these flex lines must be identical with conic (5). In fact the condition that these conics be the same for a proper curve is just that $a = -(n+1)/(n-1)$ whether n be odd or even, the equations reducing to

$$4u_1u_2 = (n-1)^2u_3^2, \text{ or } (n-1)^2x_1x_2 = x_3^2.$$

For the even case this conic is on the biflecnodal lines and we may summarize as follows:

If $a = -(n+1)/(n-1)$, *the conic on the flex lines* $t^n + 1 = 0$ *touches the curve at the points* $t^n - 1 = 0$ *(as well as at* $0, \infty$*). When n is odd the flex tangents touch curve and conic at the same points and thus count for four common lines each. When n is even the flex lines in question are biflecnodal tangents, which however do not touch the curve elsewhere. In either case all common lines of curve and conic are comprised in the flex (or biflecnodal) tangents and the tangents at the contacts.*

---

[14] Winger, "Self-projective rational curves of the fourth and fifth orders," *American Journal of Mathematics*, vol. 36 (1914), p. 75, where the curve is taken in slightly different form.

III. Metrical Specializations.

16. All of the foregoing curves can be given a metrical setting. The most effective way to do this is to introduce circular coördinates x, $\bar{x}$ by means of the relations

$$(14) \qquad \begin{aligned} x &= X + iY = x_1/x_3 \\ \bar{x} &= X - iY = x_2/x_3 \end{aligned}$$

where X, Y are ordinary Cartesian coördinates and t is to be considered as a complex number of absolute value 1.[15] The general equation of the curve now becomes[16]

$$(15) \qquad x = (t^{n+1} + at)/(t^n + 1), \qquad t = \cos\theta + i\sin\theta \equiv \operatorname{cis}\theta.$$

In this form the axes of reflexion are axes of symmetry equi-spaced about a point, the Cartesian origin. When n is even this point is a center of symmetry, i. e. a center of the curve. All of the curves are circular, touching the circular rays at I and J. The invariant conics (when proper) are concentric circles. When $a = (n+2)/(n-2)$, the self-dual case, and n is odd we get the curves treated by Duncan (l. c.).

To write the parametric equations of (15) in Cartesian coördinates, we have, expressing t in its trigonometric form

$$x = [\operatorname{cis}(n+1)\theta + a\operatorname{cis}\theta]/(1 + \operatorname{cis} n\theta).$$

Changing the denominator to half angle functions, we get

$$x = [\operatorname{cis}(n+1)\theta + a\operatorname{cis}\theta]/[2\cos(n\theta/2)\operatorname{cis}(n\theta/2)],$$

whence, dividing by cis $(n\theta/2)$,

$$x \equiv X + iY = \{\operatorname{cis}[(n+2)\theta/2] + a\operatorname{cis}[(2-n)\theta/2]\}/2\cos(n\theta/2).$$

Equating real and imaginary parts, we have the desired equations:

$$(16) \qquad \begin{aligned} 2X &= \{\cos[(n+2)\theta/2] + a\cos[(n-2)\theta/2]\}/\cos(n\theta/2) \\ 2Y &= \{\sin[(n+2)\theta/2] - a\sin[(n-2)\theta/2]\}/\cos(n\theta/2). \end{aligned}$$

[15] One equation suffices for this representation, though associated with it is the conjugate $\bar{x} = (at^n + 1)/(t^{n+1} + t)$.

[16] This is really a matter of interpretation rather than a transformation of coördinates. We should have been led to the same result had we written the equations of the curve and the generators of the group in circular coördinates to begin with and then restricted the curve to admit the group.

17. $a=-1$. Another striking metrical case is the curve for which $a=-1$. If we set $a=-1$ in (16) and apply trigonometric formulas for changing sums to products, we get

$$X=-\tan(n\theta/2)\sin\theta, \qquad Y=\tan(n\theta/2)\cos\theta.$$

Rotating the axes through 90°, we have $X=-Y'$, $Y=X'$ and the equations become, dropping primes,

$$X=\tan(n\theta/2)\cos\theta, \qquad Y=\tan(n\theta/2)\sin\theta, \tag{17}$$

whence we write down at once the polar form

$$\rho=\tan(n\theta/2). \tag{18}$$

Therefore, *if* $a=-1$ *our curves are projectively equivalent to the polar tangent curves* (18).[17]

Or, reversing the last step, we may say that *the curve* $\rho=\tan(n\theta/2)$, *where* n *is an integer (not zero), is a rational curve of order* $n+2$ *belonging to the class of curves of this paper.*

18. Again, let $a=0$. We have at once from (16), putting $(n+2)\theta/2=\phi$

$$\begin{aligned} 2X&=\sec[n\phi/(n+2)]\cos\phi \\ 2Y&=\sec[n\phi/(n+2)]\sin\phi \end{aligned} \tag{19}$$

which in polar form becomes

$$2\rho=\sec[n\phi/(n+2)]. \tag{20}$$

In this form the curve, which is obviously the inverse with respect to the pole of the rose curve $\rho=\cos[n\phi/(n+2)]$, has been widely studied under the name of the epi or Cotes's spiral.[18] Hence,

When $a=0$, *our curve may be projected into the epi* (20).

University of Washington,
Seattle, Washington.

[17] The curves $\rho=a\tan m\theta+k$, where m is rational, have been studied by Stratton, *American Mathematical Monthly*, vol. 43 (1936), p. 398 ff. He discusses the symmetry but not the implied group properties of these curves.

[18] For reference see Loria, *Spezielle Algebraische und Transzendente Ebene Kurven*, I, pp. 366, 367.

CONTRIBUTION À L'ÉTUDE DES SYSTÈMES DE CHOSES NORMÉES.*

Par V. Glivenko.

I. Position du problème.

1. Dans un article récent,[1] j'ai étudié les systèmes de choses normées, c'est à dire les ensembles S d'éléments $a, b, \cdots$ satisfaisant aux axiomes suivants:

Axiomes des choses.

1°. L'ensemble S contient des couples d'éléments a, b liés entre eux par une relation $a \subset b$ telle que $a \subset b$ et $b \subset a$ entraîne $a = b$ et inversement, et que $a \subset b$ et $b \subset c$ entraîne $a \subset c$.

2°. À tout couple d'éléments a, b de l'ensemble S correspond un élément ab, de S, tel que $ab \subset a$, $ab \subset b$ et que $x \subset a$ et $x \subset b$ entraîne $x \subset ab$.

3°. À tout couple d'éléments a, b de l'ensemble S correspond un élément $a + b$, de S, tel que $a \subset a + b$, $b \subset a + b$ et que $a \subset y$ et $b \subset y$ entraîne $a + b \subset y$.

4°. L'ensemble S contient un élément 0 tel que, quel que soit l'élément z de S, on a $0 \subset z$.

Axiome de la norme.

À tout élément a de l'ensemble S correspond un nombre non négatif $|a|$, norme de cet élément, tel que $a \subset b$ et $a \neq b$ entraîne $|a| < |b|$, qu'on a

$$|a + b| + |ab| = |a| + |b|$$

et qu'on a $|0| = 0$.

Actuellement, nous ne considérons qu'un cas particulier des dits systèmes, que nous appellerons systèmes de choses *complètement* normées et qui satisfont, par définition, aux axiomes supplémentaires suivants:

Axiome supplémentaire des choses.

5°. L'ensemble S contient un élément 1 tel que, quel que soit l'élément z de S, on a $z \subset 1$.

---

* Received January 15, 1937.

[1] "Géometrie des systèmes de choses normées," *American Journal of Mathematics*, vol. 58 (1936), pp. 799-828.

Axiome supplémentaire de la norme.

On a $|1| = 1$.

Nous appelons systèmes distributifs les systèmes où, pour tous les trois choses a, b, c a lieu la loi distributive dans la forme

$$ac + bc = (a + b)c$$

ou bien dans la forme

$$(a + c)(b + c) = ab + c$$

qui est équivalente à la précédente.

2. Dans l'article cité, j'ai introduit la notion d'espace métrique presque ordonné. C'est l'espace métrique D contenant un point que nous appelons origine et qui possède les propriétés suivantes (1′ et 2′). Rappelons que, (a, b) désignant la distance de point a et de point b, on dit qu'un point c se trouve entre deux points a et b si l'on a

$$(a, c) + (c, b) = (a, b).$$

Convenons de dire maintenant qu'un point a est plus prochain qu'un point b, ou bien que b est plus lointain que a, si a se trouve entre l'origine et b. Alors:

1′. Si les points a et b, de D, sont plus prochains qu'un point x, chaque point qui se trouve entre a et b est, lui-aussi, plus prochain que x; de même, si les points a et b sont plus lointains que y, chaque point qui se trouve entre a et b est, lui-aussi, plus lointain que y.

2′. Parmi les points, de D, qui se trouvent entre les deux points donnés quelconques, il existe un qui est le plus prochain et il existe un autre qui est le plus lointain.

J'ai établi que tout système S de choses normées est un espace métrique presque ordonné où la distance de a et de b est égale à

$$|a + b| - |ab|.$$

Nous appelons espaces transitifs les espaces métriques presque ordonnés où la condition suivante est remplie:

T. Si un point c, de D, se trouve entre x et y et si tous les deux points x et y se trouvent entre a et b, le point c se trouve, lui-aussi, entre a et b.

Convenons de dire, pour abréger, que le système S est *métrisé*, lorsqu'on a pris l'expression $|a + b| - |ab|$ pour la distance de a et de b. J'ai établi que tout système métrisé distributif S, de choses normées, est un espace métrique presque ordonné transitif.

Tous les deux résultats mentionnés, concernant les relations entre les systèmes de choses normées et les espaces métriques, ont été obtenus en prenant toujours la chose 0 du système pour l'origine de l'espace correspondant. Or, il est naturel de poser le problème s'il est possible, ou non, de prendre une autre chose $u \neq 0$ pour l'origine, en sorte qu'il y restent intactes les conditions 1′ et 2′ (dans le cas général) et la condition T (dans le cas des systèmes distributifs).

La solution de ce problème étant intéressante grâce à la simplicité des énoncés définitifs, je suis décidé à la publier ici.

II. Condition de Dedekind.

3. Dans tout ce qui suit, il nous sera d'une grande importance un fait qui a été bien éclairci, entre autres, dans les travaux de M. Garrett Birkhoff.[2] C'est que *tout système S de choses normées satisfait à la condition de Dedekind.* Cette dernière peut s'énoncer comme il suit:

(D_1) Quelles que soient les trois choses a, b, c de S, où $a \subset c$, on a

$$(a+b)c = a + bc.$$

On peut lui attribuer aussi une autre forme que voici:

(D_2) Quelles que soient les trois choses a, b, c de S, on a

$$(ac+b)c = ac + bc.$$

On établit sans peine l'équivalence de (D_1) et de (D_2) en remarquant que, pour qu'on ait $a \subset c$, il faut et il suffit qu'on ait $ac = a$.

Démontrons maintenant que tout système S de choses normées satisfait à (D_2). En premier lieu, on a toujours

$$ac + bc \subset (ac+b)c.$$

Pour s'en convaincre, il suffit de remarquer qu'on a $ac \subset ac + b$ et $ac \subset c$, par suite $ac \subset (ac+b)c$, et qu'on a $bc \subset ac + b$ et $bc \subset c$, par suite $bc \subset (ac+b)c$. Il nous reste donc à établir l'égalité

$$|ac + bc| = |(ac+b)c|.$$

On s'appuie ici sur un principe général. En effet, s'il était quelque part $x \subset y$ et $x \neq y$, il serait nécessairement $|x| < |y|$. Donc, de $x \subset y$ et $|x| = |y|$ il s'ensuit toujours $x = y$.

Pour établir l'égalité $|ac + bc| = |(ac+b)c|$, il suffit d'effectuer un simple calcul, à savoir:

---

[2] "On the combination of subalgebras" et "Applications of lattice algebra," *Proceedings of the Cambridge Philosophical Society*, vol. 23 (1933), pp. 441-469, et vol. 30 (1934), pp. 115-112 respectivement.

$$\begin{aligned}|(ac+b)c| &= |ac+b|+|c|-|ac+b+c|\\ &= |ac|+|b|-|acb|+|c|-|b+c|\\ &= |ac|-|acbc|+|bc| = |ac+bc|.\end{aligned}$$

4. Dans l'article cité, j'ai démontré que, dans l'espace formé par un système métrisé S, un point c se trouve entre deux points a et b si et seulement si l'on a, dans S,

$$(E_1) \qquad ac+bc = c = (a+c)(b+c).$$

Grâce à la condition de Dedekind, cette condition (E_1) peut se présenter aussi sous une forme distincte. En effet, en se servant de la condition de Dedekind, on obtient:

$$ac+bc = (ac+b)c \quad \text{et} \quad (a+c)(b+c) = c+(a+c)b.$$

Par conséquent, les égalités (E_1) peuvent s'écrire

$$(ac+b)c = c = c+(a+c)b,$$

ou, ce qui revient au même,

$$(E_2) \qquad (a+c)b \subset c \subset ac+b.$$

C'est la forme en question.

III. La notion d'extrémité.

5. Appelons *extrémité* d'un système S chaque chose u, de S, qui possède les deux propriétés suivantes:

1) A la chose u correspond une autre chose $\tilde{u}$, de S, telle que

$$(1) \qquad u\tilde{u} = 0,$$

$$(2) \qquad u+\tilde{u} = 1.$$

2) Quelles que soient les deux choses a et b, de S, on a

$$(3) \qquad (a+b)u = au+bu,$$

$$(4) \qquad (a+b)\tilde{u} = a\tilde{u}+b\tilde{u}.$$

Il est évident que, u étant une extrémité, $\tilde{u}$ l'est aussi; nous dirons que u et $\tilde{u}$ sont les extrémités *opposées*.

Établissons tout d'abord que, u étant une extrémité d'un système métrisé S, de choses complètement normées, on a, quelles que soient les deux choses a et b, de S,

$$ab + u = (a + u)(b + u).$$

Puisqu'on a toujours $ab + u \subset (a + u)(b + u)$, pour établir l'égalité qui vient d'être écrite, il suffit de démontrer l'égalité

$$|ab + u| = |(a + u)(b + u)|.$$

A cet effet, remarquons qu'on a

$$|(a + u)(b + u)| = |a + u| + |b + u| - |a + b + u|.$$

Or,

$$|a + u| + |b + u| = |a| + |u| - |au| + |b| + |u| - |bu|$$

et, en tenant compte de (3),

$$|a + b + u| = |a + b| + |u| - |(a + b)u|$$
$$= |a + b| + |u| - |au + bu| = |a + b| + |u| - |au| - |bu| + |abu|.$$

Donc,

$$|(a + u)(b + u)| = |a| + |b| - |a + b| + |u| - |abu|.$$

Or,

$$|a| + |b| - |a + b| = |ab|$$

et

$$|u| - |abu| = |u| - |ab| - |u| + |ab + u| = |ab + u| - |ab|.$$

Donc,

$$|(a + u)(b + u)| = |ab + u|.$$

Remarque. Le fait que u est une extrémité n'y joue, de fait, aucun rôle. Quelle que soit la chose déterminée c, de S, l'égalité

$$(a + b)c = ac + bc \tag{5}$$

a pour conséquence l'égalité

$$ab + c = (a + c)(b + c). \tag{6}$$

Nous venons de voir que ceci a lieu dans les systèmes S de choses normées. M. O. Ore[3] a réussi à établir une proposition plus générale, á savoir que l'égalité (6) est une conséquence de (5) dans tous les systèmes où la condition de Dedekind est remplie.

6. THÉORÈME I. *La chose u étant une extrémité d'un système S de choses complètement normées, son opposée $\tilde{u}$ ne peut être définie que d'une manière univoque.*

---

[3] "On the foundation of abstract algebra," I, *Annals of Mathematics*, vol. 36 (1935), p. 416.

Démonstration. La chose u étant une extremité, on a, quelles que soient les deux choses a et b,

$$(a+b)u = au + bu$$

et, comme nous l'avons vu tout à l'heure,

$$ab + u = (a+u)(b+u).$$

Prenons maintenat les choses x et y qui possèdent, tous les deux, les propriétés de l'extrémité $\tilde{u}$, de sorte qu'on a, en particulier,

$$\begin{array}{ll} ux = 0, & uy = 0, \\ u + x = 1, & u + y = 1. \end{array}$$

Il en résulte qu'on a

$$(x+y)u = xu + yu = 0, \qquad xy + u = (x+u)(y+u) = 1.$$

On en obtient, en se servant de (D_1),

$$xy = xy + (x+y)u = (xy+u)(x+y) = x+y,$$

d'où $x = y$.

THÉORÈME II. *La chose u étant une extrémité d'un système métrisé S, de choses complètement normées, la condition nécessaire et suffisante pour que, dans l'espace formé par S, un point a se trouve entre u et un autre point b, est qu'on ait*

$$ub \subset a \subset u + b.$$

Démonstration. D'après (E_2), a se trouve entre u et b si et seulement si l'on a

$$(b+a)u \subset a \subset ba + u.$$

Or, ceci équivaut à

$$bu + au \subset a \subset (b+u)(a+u),$$

ce qui equivaut, à son tour, à

$$ub \subset a \subset u + b.$$

THÉORÈME III. *Les choses u et $\tilde{u}$ étant deux extrémités opposées d'un système métrisé S, dans l'espace formé par S chaque point z se trouve entre u et $\tilde{u}$.*

Démonstration. En vertu du théorème II, pour que z se trouve entre u et $\tilde{u}$, il suffit qu'on ait

$$u\tilde{u} \subset z \subset u + \tilde{u}.$$

Or, ceci est toujours vrai, parce que $u\tilde{u} = 0$ et $u + \tilde{u} = 1$.

THÉORÈME IV. *Les choses u et ũ étant deux extrémités opposées d'un système métrisé S, si, dans l'espace formé par S, un point a se trouve entre u et b, alors b se trouve entre a et ũ et réciproquement.*

Démonstration. En vertu du théorème II, si a se trouve entre u et b, on a

$$ub \subset a \subset u + b.$$

Il s'ensuit que, d'une part,

$$\tilde{u}a \subset \tilde{u}(u + b) = \tilde{u}u + \tilde{u}b = \tilde{u}b \subset b$$

et que, d'autre part,

$$b \subset \tilde{u} + b = (\tilde{u} + u)(\tilde{u} + b) = \tilde{u} + ub \subset \tilde{u} + a.$$

En somme, on a

$$\tilde{u}a \subset b \subset \tilde{u} + a,$$

ce qui suffit, d'après le théorème II, pour que b se trouve entre a et $\tilde{u}$. La réciproque se démontre, naturellement, de même.

IV. Théorème fondamental.

7. Il serait très commode d'avoir des symboles spéciaux pour le plus prochain et le plus lointain des points que se trouvent entre les deux points donnés, a et b. Dans la suite, nous *employerons les symboles* $a \wedge b$ (pour le point le plus prochain) *et* $a \vee b$ (pour le point le plus lointain).

Ceci posé, le théorème fondamental de cet article s'énoncera comme il suit:

THÉORÈME V. *Dans l'espace formé par un système métrisé S de choses complètement normées, un point u peut être pris pour l'origine si et seulement si la chose u est une extrémité de S. Dans ce cas, on aura, en désignant par ũ l'extrémité opposée de u:*

$$(7) \quad \begin{aligned} a \wedge b &= ab + u(a + b) \\ &= ab + ua + ub = (a + b)(u + ab) = (a + b)(u + a)(u + b), \end{aligned}$$

$$(8) \quad \begin{aligned} a \vee b &= ab + \tilde{u}(a + b) \\ &= ab + \tilde{u}a + \tilde{u}b = (a + b)(\tilde{u} + ab) = (a + b)(\tilde{u} + a)(\tilde{u} + b). \end{aligned}$$

Démonstration. Nous utiliserons partout le fait que, si un point c se trouve entre deux points a et b, on a nécessairement

$$ab \subset c \subset a + b$$

(ceci est immédiat, car, si c se trouve entre a et b on a, d'après (E_1),

$$ab \subset (a + c)(b + c) = c = ac + bc \subset (a + b)$$

et que, lorsque l'un des points a ou b est une extrémité, la réciproque est vraie elle-aussi (ceci résulte du théorème II).

Soit maintenant S un système métrisé, de choses complètement normées, et soit u une extrémité de ce système. Essayons de prendre u pour l'origine de l'espace formé par S. Alors, par définition, un point a sera plus prochain qu'un autre point b, ou bien, b sera plus lointain que a, si et seulement si a se trouvera entre u et b. Nous allons voir que tous les deux conditions 1′ et 2′ y seront remplies et que les points $a \wedge b$ et $a \vee b$ y seront toujours (7) et (8).

Ad 1′. Soit c un point qui se trouve entre deux points a et b et soit x un point qui est plus prochain que a et b. Il est à montrer que x est aussi plus prochain que c. Autrement dit, soit

$$\begin{gathered} ab \subset c \subset a + b, \\ ua \subset x \subset u + a, \\ ub \subset x \subset u + b. \end{gathered}$$

Il est à montrer que

$$uc \subset x \subset u + c.$$

Or, ceci résulte des relations:

$$uc \subset u(a + b) = ua + ub \subset x \subset (u + a)(u + b) = u + ab \subset u + c.$$

Ad 2′. Soit c un point qui se trouve entre deux points a et b et soit y un point qui est plus lointain que a et b. Il est à montrer que y est aussi plus lointain que c. Autrement dit, soit

$$\begin{gathered} ab \subset c \subset a + b, \\ uy \subset a \subset u + y, \\ uy \subset b \subset u + y. \end{gathered}$$

Il est à montrer que

$$uy \subset c \subset u + y.$$

Or, ceci résulte des relations:

$$uy \subset ab \subset c \subset a + b \subset u + y.$$

Ad $a \wedge b$. Posons

$$x = ab + u(a + b) = ab + ua + ub.$$

Alors, on a, tout d'abord,

$$x = (a + b)(u + ab) = (a + b)(u + a)(u + b).$$

C'est une conséquence immédiate de (D_1). Puis, x se trouve entre a et b. Pour s'en convaincre, il suffit de remarquer que c'est (E_1) qui y est remplie, car on a, en tenant compte de (D_1),

$$ax + bx = a(ab + u(a + b)) + b(ab + u(a + b)) = ab + ua + ub = x$$

et

$$\begin{aligned}(a + x)(b + x) &= (a + (a + b)(u + ab))(b + (a + b)(u + ab)) \\ &= (a + b)(u + a)(u + b) = x.\end{aligned}$$

Soit enfin c un autre point qui se trouve entre a et b. Il est à montrer que x est plus prochain que c. Autrement dit, soit

$$ab \subset c \subset a + b.$$

Il est à montrer que

$$uc \subset x \subset u + c.$$

Or, ceci résulte des relations

$$uc \subset u(a + b) \subset ab + u(a + b) = x = (a + b)(u + ab) \subset u + ab \subset u + c.$$

Ad $a \vee b$. Posons

$$y = ab + \tilde{u}(a + b) = ab + \tilde{u}a + \tilde{u}b.$$

Alors on a, tout d'abord,

$$y = (a + b)(\tilde{u} + ab) = (a + b)(\tilde{u} + a)(\tilde{u} + b).$$

Puis, y se trouve entre a et b. Tout cela se démontre de même que les choses analogues pour x, en remplaçant x par y et u par $\tilde{u}$. Soit enfin c un autre point qui se trouve entre a et b. Il est à montrer que y est plus lointain que c. Ce dernier signifie que c se trouve entre u et y. Or, c'est équivalent, en vertu du théorème IV, à ce que y se trouve entre c et $\tilde{u}$. Autrement dit, soit

$$ab \subset c \subset a + b.$$

Il est à montrer que

$$\tilde{u}c \subset y \subset \tilde{u} + c.$$

Celui-ci se démontre de même que la chose analogue pour x, en remplaçant toujours x par y et u par $\tilde{u}$.

Nous avons démontré ainsi la suffisance de la condition du théorème.

Soit maintenant S un système métrisé, de choses complètement normées, et soit u un point, de l'espace formé par S, qui peut être pris pour l'origine de cet espace. Nous allons voir que u doit être une extrémité de S, c'est-à-dire qu'il doit posséder les propriétés 1) et 2) (n° 5).

En vertu de la propriété 2′ de l'origine, le point u ne peut être choisi pour l'origine que si, après ce choix, parmi les points qui se trouvent entre les deux points arbitraires a et b, il existera un qui sera le plus prochain et il existera un autre qui sera le plus lointain. Or, tous les points de l'espace formé par S

se trouvent entre les deux points déterminés, savoir entre 0 et 1, car on a, pour chaque chose z, de S,

$$0z + 1z = z = (0 + z)(1 + z),$$

c'est à dire que (E_1) y est remplie. Donc, on peut affirmer, en particulier, que, parmi *tous* les points de l'espace formé par S, il existera un qui sera le plus lointain. Soit v ce point. En particulier, le point v sera plus lointain que le point 0 lui-même, c'est-à-dire que 0 se trouvera entre u et v, ou bien que

$$u0 + v0 = 0 = (u + 0)(v + 0),$$

ou encore

$$uv = 0. \tag{9}$$

De même, le point v sera plus lointain que le point 1, c'est-à-dire que 1 se trouvera, lui-aussi, entre u et v, ou bien que

$$u1 + v1 = 1 = (u + 1)(v + 1),$$

ou encore

$$u + v = 1. \tag{10}$$

Revenons aux points quelconques a et b. Dire que, parmi les points c qui se trouvent entre a et b, il existera un qui est le plus prochain, c'est dire qu'il existera un point x tel que, premièrement, x se trouve entre a et b et, deuxièmement, quel que soit le point c se trouvant entre a et b, x se trouve entre u et c. Alors, puisque les points ab et $a + b$ se trouvent toujours entre a et b, il existera, en particulier, un point x qui se trouve, en même temps, entre a et b, entre u et ab et entre u et $a + b$, de sorte qu'on a

$$ab \subset x \subset a + b,$$
$$uab \subset x \subset u + ab,$$
$$u(a + b) \subset x \subset u + (a + b).$$

Nous allons voir, tout d'abord, que le seul point x pouvant satisfaire à ces trois conditions est

$$x = ab + u(a + b) = (a + b)(u + ab).$$

En effet, on doit y avoir

$$ab + u(a + b) \subset x \subset (a + b)(u + ab).$$

Par suite, dire qu'il existera un point qui se trouve entre a et b et qui satisfait à trois conditions en question, c'est dire, en particulier, que le point $ab + u(a + b)$ se trouvera entre a et b. Or, dans l'article cité, j'ai établi [4] que, lorsqu'on a

[4] *Au cours de la démonstration du théorème VI*, pp. 821-823.

$$(a+b)u \neq au + bu,$$

alors le point $ab + u(a+b)$ ne se trouve pas entre a et b. Par suite, dire que le point $ab + u(a+b)$ se trouvera entre a et b, c'est dire, en particulier, qu'on aura

$$(a+b)u = au + bu. \tag{11}$$

Pareillement, dire que, parmi les points c qui se trouvent entre a et b, il existera un qui est le plus lointain, c'est-dire qu'il existera un point y tel que, premièrement, y se trouve entre a et b et, deuxièmement, quelque soit le point c se trouvant entre a et b, c se trouve aussi entre u et y. Or, ce dernier revient au dire que y se trouve entre c et v. En effet le point v étant plus lointain que tout autre point, les points c et y se trouvent, tous les deux, entre u et v, de sorte qu'on a

$$(u, c) + (c, v) = (u, v),$$
$$(u, y) + (y, v) = (u, v).$$

On en tire:

$$(u, c) + (c, v) = (u, y) + (y, v).$$

De plus, lorsque c se trouve entre u et y, on a

$$(u, c) + (c, y) = (u, y).$$

En comparant, on obtient:

$$(c, y) + (y, v) = (c, v).$$

Cela établi, on peut affirmer qu'on aura:

$$(a+b)v = av + bv. \tag{12}$$

On la démontre de même que la chose analogue pour x, en remplaçant x par y et u par v.

En résumé, on voit qu'à la chose u doit correspondre une chose v, de sorte que u et v satisfassent aux égalités (9), (10), (11) et (12). Or, ces dernières expriment précisement les propriétés 1) et 2) caractérisant l'extrémité u et son opposée $\tilde{u}$.

Nous avons démontré ainsi la nécessité de la condition du théorème.

8. Théorème VI. *Si, dans l'espace formé par un système métrisé S, de choses complètement normées, tout point de l'espace peut être pris pour l'origine, le système S est nécessairement distributif.*

C'est un simple corollaire du Théorème V. En effet, lorsque tout point c, de l'espace formé par S, peut être pris pour l'origine, alors, en vertu du

théorème V, toute chose c, de S, est une extrémité de S. Dans ce cas, quelles que soient les choses a, b, c, de S, on doit avoir

$$(a+b)c = ac + bc.$$

C'est la loi distributive.

V. Cas des systèmes distributifs.

9. La condition T, nécessaire et suffisante pour que l'espace métrique presque ordonné soit formé par un système distributif, reste intacte quand on varie le choix de l'origine de cet espace. En effet, la condition T ne contient aucune indication, soit explicite soit implicite, sur le choix de l'origine.

Il en est autrement pour une autre condition, T bis, établie dans mon article cité. Pour bien comprendre tout ce qui suit, rappelons d'abord un théorème auxiliare qui y était établi, à savoir:

Lorsque, dans l'espace métrique formé par un système de choses normées, on prend 0 *pour l'origine, un point a est plus prochain qu'un point b si et seulement si l'on a*

$$a \subset b.$$

Après avoir démontré ce théorème, j'ai commencé d'employer le symbole $a \subset b$, non pas seulement dans son sens primitif, celui d'une relation entre les deux choses a et b d'un système, mais aussi pour exprimer que le point a, de l'espace formé sur ce système, est plus prochain que le point b. Pareillement, il y était établi que, lorsqu'on prend 0 pour l'origine, on a $a \wedge b = ab$ et $a \vee b = a + b$. Après ceci, j'ai commencé d'employer les symboles ab et $a+b$ pour désigner les points $a \wedge b$ et $a \vee b$. C'est ce qu'on doit avoir en vue dans l'énoncé du théorème concernant la condition T bis, à savoir:

La condition nécessaire et suffisante pour qu'un espace métrique presque ordonné D soit transitif est qu'il possède la propriété suivante:

T bis. *Un point c, de D, se trouve entre deux points a et b si et seulement si l'on a*

$$ab \subset c \subset a + b.$$

On y suppose, comme je le faisais partout dans l'article cité, que c'est 0 qui y est pris pour l'origine. Donc, en tenant compte des remarques qui viennent d'être faites, l'expression $ab \subset c \subset a+b$ qui figure dans l'énoncé de ce dernier théorème peut s'interpréter de deux façons différentes. Nous allons exprimer cela explicitement en forme de deux théorèmes, VII et VIII. Ce qui est intéressant, c'est que ces théorèmes-ci restent valables avec le choix arbitraire de l'origine.

THÉORÈME VII. *La condition nécessaire et suffisante pour qu'un espace métrique presque ordonné D formé par un système S de choses normées, soit transitif est qu'il possède la propriété suivante:*

T bis. *Un point c, de D, se trouve entre deux points a et b si et seulement si l'on a, pour les choses a, b, c de S,*

$$ab \subset c \subset a + b.$$

La démonstration y est inutile, car c'est précisemment ce qui a été demontré dans mon article cité.

THÉORÈME VIII. *La condition nécessaire et suffisante pour qu'un espace métrique presque ordonné D soit transitif est qu'il possède la propriété suivante:*

T ter. *Un point c, de D, se trouve entre deux points a et b si et seulement si ce point c est plus lointain que $a \wedge b$ et plus prochain que $a \vee b$.*

Démonstration. Supposons d'abord que l'espace en question est transitif, c'est-à-dire que la condition T y est remplie, et soit c un point qui est plus lointain que $a \wedge b$ et plus prochain que $a \vee b$. Nous allons voir que c se trouve alors entre $a \wedge b$ et $a \vee b$. En effet, d'après les suppositions faites, le point $a \wedge b$ se trouve entre l'origine u et le point c, tandis que le point c se trouve entre l'origine u et le point $a \vee b$. Ce dernier peut s'exprimer aussi, en vertu du Théorème IV, en disant que le point $a \vee b$ se trouve entre le point c et le point $\tilde{u}$ qui est l'extrémité opposée de u. On a donc, en vertu du Théorème II,

$$uc \subset a \quad b \subset u + c,$$

$$\tilde{u}c \subset a \vee b \subset \tilde{u} + c.$$

De plus, comme, en vertu du Théorème III, *tous* les points se trouvent entre u et $\tilde{u}$, on a, d'après (E_1),

$$uc + \tilde{u}c = c = (u + c)(\tilde{u} + c).$$

On en déduit sans peine qu'on a

$$(a \wedge b + c)(a \vee b) \subset (u + c)(\tilde{u} + c) = c = uc + \tilde{u}c \subset (a \wedge b)c + (a \vee b).$$

En tenant compte de (E_2), ceci suffit pour affirmer que le point c se trouve entre les points $a \wedge b$ et $a \vee b$. Or, ces derniers points se trouvent entre a et b. Par suite, d'après T, le point c se trouve, lui-aussi, entre a et b. On voit ainsi que, si le point c est plus lointain que $a \wedge b$ et plus prochain que $a \vee b$, ce point c se trouve entre a et b. La réciproque étant manifeste, cela signifie que la condition T ter est remplie.

Inversement, supposons que la condition T ter soit remplie, que c se trouve entre x et y et que x et y se trouvent entre a et b. Alors, nous savons qu'on a

$$xy \subset c \subset x + y,$$
$$ab \subset x \subset a + b,$$
$$ab \subset y \subset a + b,$$

d'où

$$ab \subset c \subset a + b.$$

Ceci posé, on peut démontrer que le point c est plus lointain que $a \wedge b$ et plus prochain que $a \vee b$, c'est à dire que

$$uc \subset a \wedge b \subset u + c,$$
$$\bar{u}c \subset a \vee b \subset \bar{u} + c,$$

ou bien que

$$uc \subset ab + u(a + b) \subset u + c,$$
$$\bar{u}c \subset ab + \bar{u}(a + b) \subset \bar{u} + c.$$

Pour s'en convaincre, il suffit de remarquer que, lorsqu'on a $ab \subset c$, on a aussi

$$ab + u(a + b) \subset c + u(a + b) \subset c + u,$$
$$ab + \bar{u}(a + b) \subset c + \bar{u}(a + b) \subset c + \bar{u},$$

et que, lorsqu'on a $c \subset a + b$, on a aussi

$$uc \subset u(a + b) \subset ab + u(a + b),$$
$$\bar{u}c \subset \bar{u}(a + b) \subset ab + \bar{u}(a + b).$$

Ceci établi, il en resulte, d'après T ter, que c se trouve entre a et b. On voit ainsi que, si le point c se trouve entre x et y et que x et y se trouvent entre a et b, le point c se trouve, lui-aussi, entre a et b. C'est la condition T, de sorte que l'espace est transitif.

10. Dans le cas des systèmes distributifs, l'énoncé du théorème fondamental se simplifie comme il suit.

THÉORÈME IX. *Dans l'espace formé par un système métrisé distributif S, de choses complètement normées, un point u peut être pris pour l'origine si et seulement si la chose u, de S, possède le complément, c'est à dire qu'il existe la chose ū liée avec u par les relations*

$$u\bar{u} = 0,$$
$$u + \bar{u} = 1.$$

C'est une conséquence immédiate du Théorème V. En effet, la propriété de u qui figure dans l'énoncé du présent théorème n'est autre chose que la

propriété 1) de l'extrémité. Quant à la propriété 2) de l'extrémité, elle est remplie, dans les systèmes distributifs, de soi-même.

11. Il est intéressant de considérer encore le cas particulier des systèmes *ordonnés*, c'est-à-dire des systèmes où, pour chaque couple de choses a, b, on a nécessairement, ou bien $a \subset b$ ou bien $b \subset a$.

THÉORÈME X. *Dans l'espace formé par un système métrisé ordonné S, de choses complètement normées, les seuls points qui peuvent être pris pour l'origine sont les point* 0 *et* 1.

En effet, l'origine devant être, d'après le théorème V, une extrémité de S, soit u cette extrémité et soit $\tilde{u}$ son opposée. Alors, si l'on a $u \subset \tilde{u}$, on a $u = u\tilde{u} = 0$. Si l'on a, au contraire, $\tilde{u} \subset u$, on a $u = u + \tilde{u} = 1$.

VI. Un exemple.

12. Terminons par un exemple. Soit S l'ensemble dont les éléments $a, b, \cdots$ sont les fonctions mesurables $\phi(\xi)$, d'une variable réelle ξ, définies dans l'intervalle $0 \leq \xi \leq 1$ et telles que $0 \leq \phi(\xi) \leq 1$. Convenons d'écrire $a = b$ si les fonctions correspondantes ne diffèrent l'une de l'autre que sur un ensemble de mesure nulle, et posons $a \subset b$ si, pour les fonctions correspondantes $\phi(\xi)$ et $\psi(\xi)$, on a presque partout $\phi(\xi) \leq \psi(\xi)$. Enfin, a étant la fonction $\phi(\xi)$, posons

$$| a | = \int_0^1 \phi(\xi)\, d\xi.$$

Cela posé, l'ensemble S est un système distributif de choses complètement normées. Ici, on a, a et b étant respectivement les fonctions $\phi(\xi)$ et $\psi(\xi)$,

$$ab = \min[\phi(\xi), \psi(\xi)], \qquad a + b = \max[\phi(\xi), \psi(\xi)].$$

Quant aux extrémités opposées, u et $\tilde{u}$, ce sont respectivement les fonctions $\omega(\xi)$ et $\tilde{\omega}(\xi)$ de la forme suivante. Soient Ξ_1 et Ξ_2 deux ensembles mesurables, de points ξ, dont la partie commune est vide et dont la somme est l'intervalle $0 \leq \xi \leq 1$, d'ailleurs arbitraires. Alors

$$\omega(\xi) = \begin{cases} 0 \text{ sur } \Xi_1, \\ 1 \text{ sur } \Xi_2, \end{cases} \qquad \tilde{\omega}(\xi) = \begin{cases} 1 \text{ sur } \Xi_1, \\ 0 \text{ sur } \Xi_2. \end{cases}$$

Dans l'espace formé par le système S métrisé, la distance de point a et de point b, qui sont respectivement les fonctions $\phi(\xi)$ et $\psi(\xi)$, est donnée par la formule

$$(a, b) = \int_0^1 | \phi(\xi) - \psi(\xi) |\, d\xi.$$

Ici, un point c, qui est la fonction $\chi(\xi)$, se trouve entre deux points a et b, qui sont respectivement les fonctions $\phi(\xi)$ et $\psi(\xi)$, si et seulement si l'on a presque partout

$$\min[\phi(\xi), \psi(\xi)] \leq \chi(\xi) \leq \max[\phi(\xi), \psi(\xi)].$$

Pour l'origine, on peut y prendre un point u qui est l'une des fonctions $\omega(\xi)$ définies plus haut. Enfin, a et b étant respectivement les fonctions $\phi(\xi)$ et $\psi(\xi)$, on a

$$a \wedge b = \begin{cases} \min\,[\phi(\xi), \psi(\xi)] & \text{sur } \Xi_1, \\ \max\,[\phi(\xi), \psi(\xi)] & \text{sur } \Xi_2, \end{cases}$$

$$a \vee b = \begin{cases} \max\,[\phi(\xi), \psi(\xi)] & \text{sur } \Xi_1, \\ \min\,[\phi(\xi), \psi(\xi)] & \text{sur } \Xi_2. \end{cases}$$

Appendice.

Je profite de l'occasion pour corriger une petite erreur admise dans mon article cité. On lit, p. 804: " Pour avoir un exemple d'espace métrique presque ordonné, prenons un ensemble arbitraire de nombres réels, où $(a, b) = |a - b|$. Il est aisé de voir que le rôle de l'origine peut être joué par un nombre quelconque appartenant à cet ensemble." Le Théorème X du présent article nous apprend que cet exemple est incorrect.

Cette même erreur se répète p. 818, où l'on lit: " Pour avoir un exemple d'espace métrique presque ordonné transitif, il suffit de rappeler l'exemple déjà cité d'un ensemble arbitraire de nombres réels, où $(x, y) = |x - y|$."

Moscou, U. S. S. R.

TOPOLOGICAL PROPERTIES OF THE SOLUTIONS OF ORDINARY DIFFERENTIAL EQUATIONS.*

By T. M. Cherry.

The conception that any solution of a set of differential equations possesses topological properties rests upon its representation as a curve in a suitably chosen space, and the observation that the points of this curve do not in general form a closed set. Hence we are led to consider its derived set, which consists of the curve itself together with two other sets, its ω- and α-limit sets, which are derived respectively from the "positive and negative ends" of the curve (Birkhoff [1]). The present paper rests upon the idea that solutions may be classified according to their relations with the associated ω- and α-limit sets. This idea is not new, the case of a solution which is contained by both its limit sets having been long recognized as of importance (Hadamard [1], Birkhoff and Smith [1]); but it does not appear hitherto to have been systematically developed. Such a development is found in §§ 4, 5, 7 below, in relation to which §§ 1-3 contain the necessary preliminaries and § 6 is a digression. The recurrent solutions of Birkhoff appear in § 5 as a sub-class of those I call pervasive, while the theory of general pervasive solutions in § 7 amounts to a generalization of the descriptive theory of transitive systems. In §§ 8-11 there is developed a method of constructing transitive systems which do not consist simply of a single minimal set, and yet are soluble by quadratures; the systems constructed are not however of Hamiltonian type.

The extent of previous explorations in the field with which we are concerned makes it inevitable that the present work should be distinctive in its point of view rather than its method or final results. Theorems II, IV, V and part of Theorem VI are apparently new, while the proofs of Theorems III (i), VI, VII differ more or less in detail from those given by Birkhoff. The examples in §§ 9-11 are new.

[*Note added in proof.* A Referee has drawn my attention to a recent paper of H. Hilmy [1], the ideas of which are to some extent common with those here used, though differently introduced. Hilmy's *exceptional undecomposable set* is in my terminology (§ 4) an asymptotic trajectory together with its ω- and α-limit sets; his *quasi-minimal set* is my *stratum* (§ 7); and

* Received March 4, 1937.

his example of a quasi-minimal set which is not minimal anticipates the principle used in § 9.]

1. Terminology. Let

$$dx_i/dt = X_i(x_1, \cdots, x_n), \qquad (i = 1, \cdots, n) \tag{1}$$

be a system of differential equations in which $X_1, \cdots, X_n$ are real single-valued functions of n real dependent variables $x_1, \cdots, x_n$. The conception that is fundamental to our object is that any set of real values $(x_1, \cdots, x_n)$ of these variables is a single entity, a *point* of an n-dimensional space; so that by suitably defining one or both of the terms *neighbourhood* and *distance* for this space we can utilize the topological ideas of limit-point, etc. It is customary to take the space to be Euclidean, but that this convention is unnecessarily restrictive is evident since the results to be proved are purely topological. We shall take as our fundamental definition: *A neighbourhood of a point* (a_i) *is the aggregate of all points* (x_i) *for which, for some positive number* ϵ,

$$| x_i - a_i | < \epsilon \qquad (i = 1, \cdots, n).$$

More specifically, we call this aggregate *the* ϵ*-neighbourhood of* (a_i). We utilize also *closed neighbourhoods* defined by inequalities $| x_i - a_i | \leq \epsilon$.

For convenience of language we abridge certain statements about neighbourhoods by use of the term *distance*, with the definition: *The distance between the points* (x_i), (a_i) *is the greatest of the n numbers* $| x_i - a_i |$. The symmetry and triangle axioms are thereby satisfied; and the statement "the distance between (x_i), (a_i) is equal to (less than, greater than) ϵ" is equivalent to "each of the points (x_i), (a_i) lies on the boundary of (inside, outside) the ϵ-neighbourhood of the other."

We describe the functions X_i as *regular* at a point (a_i) if *they all have continuous first derivatives in some neighbourhood of the point.* If the X_i are regular at (a_i) a solution

$$x_i = f_i(t) \qquad (i = 1, \cdots, n) \tag{2}$$

of the system (1), for which $f_i(t_0) = a_i$, is uniquely defined in some interval $| t - t_0 | < \tau$; and the solution may be continued over successive intervals overlapping with this as long as the point $(f_i(t))$ does not approach arbitrarily near to a point where the X_i are not all regular.

We describe the solution (2) as *regular* if *the* $f_i(t)$ *may be continued over all real values of* t, *and if there is a positive constant* δ *such that the* X_i *are regular in the* δ*-neighbourhood of each of its points.* We describe the solution

as *bounded* if *the* $f_i(t)$ *are bounded.* The existence of regular bounded solutions [1] is assured if the system (1) possesses an integral

$$F(x_1, \cdots, x_n) = \text{constant} = c$$

which, for some range of c, represents in our space a family of hyper-surfaces possessing closed sheets, at all points of which the X_i are regular. Our object being to investigate properties of a solution which follow merely from the hypothesis that it is bounded and regular, it would be beside the point here to develop other criteria for the existence of such solutions.

Any solution (2) specifies in our space a variable point $(f_i(t))$ which we call a *current point* and denote by a symbol such as $P(t)$; definite positions of it (fixed points of space) are denoted by symbols such as $P(t_0), P_0, P$; the aggregate of all its positions for $-\infty < t < \infty$ is a solution-curve or *trajectory.* Any *solution* specifies a trajectory together with a correspondence between its points and values of t. It is often (e. g. in the phrase "a point of a solution [trajectory]") but not always a matter of indifference whether we speak of "solution" or "trajectory"; and we use the same symbol, such as (P), for both. We call t the time-variable, and use terms such as *time, instant, prior* in connection therewith; and we can hence think of a solution as being or specifying a *motion* of a current point.

From the solution (2) we can derive another

$$x_i = f_i(t + \tau) \qquad (i = 1, \cdots, n) \tag{3}$$

in which τ is any constant. To the solutions (2), (3) correspond different current points $P(t), P(t + \tau)$ which run along the same trajectory, the former passing through any position a time τ in advance of the latter. Thus given any two positions P_0, P_1 on the trajectory, all current points thereon take the same time, T say, to pass from P_0 to P_1, and the time $(-T)$ to pass from P_1 to P_0. We call P_1 the *T-transform* [2] of P_0, and P_0 the $(-T)$-transform of P_1.

2. The auxiliary theorem. Let $\mathfrak{D}$ be a closed connected set of points, and $\mathfrak{D}'$ the (open) domain which is the sum of the δ-neighbourhoods of all points of $\mathfrak{D}$; δ being a positive constant. Let the X_i be regular at all points of $\mathfrak{D}'$. Then, from the fundamental theorems on the existence of solutions and their continuity with respect to the initial values, there follows:

---

[1] Birkhoff [1] calls such solutions *positively and negatively stable.*

[2] Poincaré [1] calls P_1 a *consequent* or *antecedent* of P_0 according as T is positive or negative.

(i) *There is a positive constant* τ *such that any current point which is in* $\mathfrak{D}$ *for* $t = t_0$ *remains within* $\mathfrak{D}'$ *throughout the interval* $|t - t_0| < \tau$ *at least.*

(ii) *Let a current point* $P(t)$ *be at* $P(t_0)$ *in* $\mathfrak{D}$ *for* $t = t_0$, *and remain within* $\mathfrak{D}'$ *for* $t_1 < t < t_2$, *(where* t_1 *may be* $-\infty$ *and* t_2 *may be* ∞*); let* ϵ *be any positive number and* T *any (finite) number between* t_1 *and* t_2*; then there exists a positive number* η, $\equiv \eta(\epsilon, T)$, *such that all current points whose distance from* $P(t_0)$ *at* $t = t_0$ *does not exceed* η *are at a distance from* $P(t)$ *not exceeding* ϵ *for all* t *in the closed interval* (t_0, T).

Following Birkhoff [1], we call (ii) the Auxiliary Theorem.

Corollary. *If* $P(t_0)$ *is a limit-point (or the sole limit-point) of a sequence of points, then* $P(t_0 + T)$ *is a limit-point (or the sole limit-point) of the T-transform of this sequence, provided* $t_1 < t_0 + T < t_2$.

3. The ω- and α-limit sets of a bounded regular solution. The following definitions and theorems are due, in their precise form, to Birkhoff [1].

Let (P) be a bounded regular solution and $P(t)$ the associated current point. Let $\{t_i\}$ be any increasing unbounded sequence. The hypothesis that the solution is bounded implies that the points $P(t_i)$ all lie in a bounded portion of the space; hence the set $\{P(t_i)\}$ has at least one limit point, called an *ω-limit point of* (P). The set of all such points is called the *ω-limit set of* (P), and denoted by $L_\omega(P)$.

The set of points similarly defined by decreasing unbounded sequences $\{t_i\}$ is called the *α-limit set of* (P), and denoted by $L_\alpha(P)$.

It is apparent that all solutions which, like (2) and (3), belong to the same trajectory define the same ω- and α-limit sets.

The derived set of (P), regarded simply as a set of points, consists of $L_\omega(P)$, $L_\alpha(P)$ together with the set (P) itself, which may or may not belong to either or both of $L_\omega(P)$, $L_\alpha(P)$.

Theorem I. *The following results are stated for the ω-limit set; there are corresponding results for the α-limit set.*

(i) *The set* $L_\omega(P)$ *is closed,* since in an arbitrary neighbourhood of any point of its derived set there are points of $L_\omega(P)$ and hence points $P(t)$ for which t is arbitrarily large.

(ii) *The set* $L_\omega(P)$ *consists of a set of bounded regular trajectories.* If Q is any point of $L_\omega(P)$ we are to prove that for any T the T-transform of Q belongs to $L_\omega(P)$.

Call $\mathfrak{D}'$ the δ-neighbourhood of the trajectory (P), i. e. the sum of the δ-neighbourhoods of all its points: δ being the constant involved by definition (§ 1) in the hypothesis that (P) is regular. By the Auxiliary Theorem, if Q_1, the T-transform of Q, belongs to $\mathfrak{D}'$, and if Q is a limit-point of the sequence $\{P(t_i)\}$ in which $t_i \to \infty$ as $i \to \infty$, then Q_1 is a limit-point of the sequence $\{P(t_i + T)\}$, and is an ω-limit-point since $t_i + T \to \infty$ as $i \to \infty$. Hence *if the T-transform of Q belongs to $\mathfrak{D}'$ it belongs to $L_\omega(P)$.*

But by § 2 (i) there is a positive constant τ such that, if $Q(t')$ belongs to $L_\omega(P)$, (every point of which is within $\mathfrak{D}'$ at a distance at least δ from its boundary), then $Q(t)$ belongs to $\mathfrak{D}'$ for $|t - t'| < \tau$; so by the result just italicized, if $Q(t)$ belongs to $L_\omega(P)$ for $t' \leqq t \leqq t''$, it belongs to $L_\omega(P)$ for $t' - \tau < t < t \ + \tau$. Hence $Q(t)$ belongs to $L_\omega(P)$ for all real t; which proves the result.

It is clearly appropriate to say that $L_\omega(P)$ consists of a closed set of trajectories or solutions.

(iii) *Corresponding to any positive number ϵ there exists a (finite) number $T, \equiv T(\epsilon)$, such that for $t > T$, $P(t)$ lies within the ϵ-neighbourhood of $L_\omega(P)$.* This is proved by reductio ad absurdum. In virtue of the result we can say that, if (P) is not a part of $L_\omega(P)$, $P(t)$ is *asymptotic* to $L_\omega(P)$ as $t \to \infty$.

(iv) From (iii) it follows that *the set $L_\omega(P)$ is connected*; and hence from (i) *it is perfect.*

(v) *To any numbers t_0 and ϵ, $(\epsilon > 0)$, corresponds a (finite) number $T', \equiv T'(\epsilon, t_0)$, greater than t_0 such that, for some t between t_0 and T', $P(t)$ is within the ϵ-neighbourhood of each point of $L_\omega(P)$*; i. e. $P(t)$ approaches its limit-points $L_\omega(P)$ uniformly. This is proved by reductio ad absurdum.

4. Classification of bounded regular solutions. It follows from § 3 (ii) that, if one point of the solution (P) belongs to $L_\omega(P)$, then all its points belong. Hence there are two mutually exclusive possibilities, which we denote by the symbols V_ω, A_ω:

V_ω : every point of (P) belongs to $L_\omega(P)$,
A_ω : no point of (P) belongs to $L_\omega(P)$.

Similarly there are two mutually exclusive possibilities, which are shewn by particular examples (§§ 9, 10 below) to be independent of the former:

V_α : every point of (P) belongs to $L_\alpha(P)$,
A_α : no point of (P) belongs to $L_\alpha(P)$.

By combination of these we classify bounded regular solutions (P) into three types, as follows:

PERVASIVE SOLUTIONS $(V_\omega V_a)$. *(P) belongs to both $L_\omega(P)$ and $L_a(P)$.*

ASYMPTOTIC SOLUTIONS $(A_\omega A_a)$. *(P) belongs to neither $L_\omega(P)$ nor $L_a(P)$.*

PERVASIVE-ASYMPTOTIC SOLUTIONS $(V_\omega A_a$ OR $V_a A_\omega)$. *(P) belongs to one but not both of $L_\omega(P)$, $L_a(P)$.*

Any solution which possesses the property V_ω will be called *positively semi-pervasive*; and similarly for the properties V_a, A_ω, A_a. Thus a solution of type $V_a A_\omega$ is negatively semi-pervasive and positively semi-asymptotic.

It is apparent that all solutions which differ merely by constants additive with t are of the same type; and the trajectory which carries a pervasive (asymptotic, etc.) solution may be described as pervasive (asymptotic, etc.).

If a solution (P) is positively semi-pervasive the set of points $L_\omega(P)$ will be called the associated *stratum*; [3] and similarly for a negatively semi-pervasive solution. If the solution is pervasive it will be shewn below (Theorem II) that the sets $L_\omega(P)$, $L_a(P)$ coincide, so that there is only one stratum associated with the solution.

A "pervasive" solution is so named because its points are everywhere dense in, or pervade, the associated stratum. The term "recurrent" would be equally appropriate, but is already standardized as denoting a sub-class of pervasive solutions, namely, uniformly pervasive solutions (§ 5 below). The terms stable-à-la-Poisson (Poincaré [2]) quasi-periodic (Husson [1]) and quasi-recurrent (Birkhoff and Smith [1]) have also been used with the same sense. An "asymptotic" solution is named from the property pointed out in § 3 (iii) (*cf.* Hadamard [1]).

In the following work the symbol $\subset$ is used to denote the relation of *strict* inclusion between the two sets; so that $S_1 \subset S_2$ excludes the possibility $S_1 = S_2$.

THEOREM II. *For a pervasive solution (P), $L_\omega(P) = L_a(P)$. For a pervasive-asymptotic solution of type $V_\omega A_a$, $L_a(P) \subset L_\omega(P)$. For a pervasive-asymptotic solution of type $V_a A_\omega$, $L_\omega(P) \subset L_a(P)$.*

Let us assume the hypothesis V_ω, i. e. that (P) is positively semi-pervasive. Every α-limit point of (P) is a limit point of a sequence $\{P(t_i)\}$ in which

[3] Hadamard [1] calls this the *domain of* (P); as this usage of the word "domain" is not in accord with its accepted modern connotation I introduce a new term.

$\{t_i\}$ decreases to $-\infty$. Since all the points of this sequence belong to (P), and hence by hypothesis to $L_\omega(P)$, and the latter set is closed, every α-limit point of (P) belongs to $L_\omega(P)$. Hence

$$\text{(4)} \qquad \textit{if } (P) \subseteqq L_\omega(P), \textit{ then } L_\alpha(P) \subseteqq L_\omega(P).$$

Similarly, *if* $(P) \subseteqq L_\alpha(P)$, *then* $L_\omega(P) \subseteqq L_\alpha(P)$.

Hence if (P) is pervasive, so that $(P) \subseteqq L_\omega(P)$ and $(P) \subseteqq L_\alpha(P)$, then $L_\alpha(P) = L_\omega(P)$.

Conversely, assume that $(P) \subseteqq L_\omega(P)$ and $L_\alpha(P) = L_\omega(P)$; then $(P) \subseteqq L_\alpha(P)$, and the solution (P) is pervasive.

Hence if $(P) \subseteqq L_\omega(P)$ and (P) is not pervasive, so that though positively semi-pervasive it is negatively semi-asymptotic, the possibility $L_\omega(P) = L_\alpha(P)$ is excluded, and from (4) there follows $L_\alpha(P) \subset L_\omega(P)$. Q. E. D.

It may be noted that for a solution of type $V_\omega A_\alpha$ we have $(P) \subset L_\omega(P)$ and not $(P) = L_\omega(P)$; for the latter alternative in combination with the Theorem just proved would give $L_\alpha(P) \subset (P)$, implying that some but not all points of (P) are α-limit points. If $(P) \subset L_\omega(P)$ the latter set consists of a non-enumerable set of trajectories, and if $(P) = L_\omega(P)$, (P) is periodic (Birkhoff [1]).

For an asymptotic solution there are no necessary relations of inclusion or exclusion between $L_\omega(P)$, $L_\alpha(P)$. Examples may be given of each of the cases (i) $L_\alpha(P) = L_\omega(P)$, (ii) $L_\omega(P)$, $L_\alpha(P)$ have no common point, (iii) one of $L_\omega(P)$, $L_\alpha(P)$ is a proper part of the other, (iv) $L_\omega(P)$, $L_\alpha(P)$ have common points, but neither contains the other.

The classical work of Poincaré [2] has often been quoted as proving, for the case in which there is a positive integral-invariant and all solutions which pass through a certain neighbourhood are bounded and regular, the existence of solutions of the types here called semi-pervasive and pervasive. This is not so; for Poincaré, taking an arbitrarily small but determinate (fixed) region $\mathfrak{D}$, proves the existence of solutions which traverse it infinitely often, and all that this shews is that the ω-limit set (or α-limit set, or both) traverse $\mathfrak{D}$. Poincaré's method is easily modified so as to yield the existence of pervasive solutions; but this is unnecessary since Carethéodory [1], who quotes Poincaré in the above incorrect sense, has proved that *the set of trajectories which are not positively (negatively) semi-pervasive is of measure zero,* and the complement of the sum of these two nul sets is the set of pervasive trajectories. Thus, with the hypothesis stated, *almost all trajectories are pervasive.*

5. Uniformly and non-uniformly pervasive solutions. The notion of a *uniformly pervasive* solution is due to Birkhoff [1], who calls it *recurrent*. From the present point of view it seems most natural to adopt as a definition the necessary and sufficient property of Birkhoff's Theorem III, instead of the notion of a minimal set, as follows:

Let (P) be a pervasive solution, and denote the associated stratum, i. e. the coincident sets $L_\omega(P)$, $L_a(P)$, by $\mathcal{S}$. Let us say that $P(t)$, starting from the position $P(t_0)$, makes an ϵ-circuit of $\mathcal{S}$ in time T, $(T > 0)$, if for some t between t_0 and $t_0 + T$ (inclusive) it is in the closed ϵ-neighbourhood of each point of $\mathcal{S}$. Then we define:

The pervasive solution (P) is uniformly pervasive or recurrent if for every positive number ϵ there is a (finite) time T, $\equiv T(\epsilon)$, within which a current point $P(t)$ makes an ϵ-circuit of $\mathcal{S}$, whatever its starting point on (P).

THEOREM III.[4] (i) *If the solution (P) is uniformly pervasive its stratum $\mathcal{S}$ consists of an aggregate of solutions each of which is uniformly pervasive with $\mathcal{S}$ as its stratum.*

(ii) *If the solution (P) is non-uniformly pervasive, every solution which has one point in the stratum $\mathcal{S}$ associated with (P) lies in $\mathcal{S}$, as do its ω- and α-limit sets; and $\mathcal{S}$ contains solutions for which both the ω- and α-limit sets are proper parts thereof.*

(i) Let any positive number ϵ be assigned. Choose any positive number ϵ' less that ϵ, let $\kappa = \epsilon - \epsilon'$, and let $T(\epsilon')$ be the time not exceeded, by definition, in any ϵ'-circuit of $\mathcal{S}$ by $P(t)$.

Let Q be any point of $\mathcal{S}$, $Q(t)$ the current point that is at Q at $t = 0$, and $R(t)$ any other current point. By the Auxiliary Theorem there is a positive η such that the distance $\overline{Q(t)R(t)}$ does not exceed κ for $0 \leq t \leq T(\epsilon')$ provided only the initial distance $\overline{QR(0)}$ does not exceed η. Take for $R(0)$ a point $P(t')$ whose distance from Q does not exceed η, a suitable choice of t' being possible since Q belongs to $\mathcal{S}$. Then for $0 \leq t \leq T(\epsilon')$ the distance $\overline{Q(t)P(t' + t)}$ does not exceed κ. But in this time-interval $P(t' + t)$ passes within a distance not exceeding ϵ' of each point of $\mathcal{S}$; so $Q(t)$ must pass within a distance not exceeding $\kappa + \epsilon'$, $= \epsilon$; and as ϵ is arbitrary, every point of $\mathcal{S}$ belongs to $L_\omega(Q)$.

But since $(Q) \subseteqq \mathcal{S}$ and the latter set is closed, $L_\omega(Q) \subseteqq \mathcal{S}$. Hence $L_\omega(Q) = \mathcal{S}$. By a similar argument we prove $L_a(Q) = \mathcal{S}$; and since the

[4] This is identical with Birkhoff [1], Theorem III. The proof of (i) differs from Birkhoff's.

time for an ϵ-circuit by $Q(t)$ has been shewn not to exceed the constant $T(\epsilon')$, whatever its starting point, (Q) is uniformly pervasive.

(ii) The hypothesis that (P) is non-uniformly pervasive implies that, for some fixed positive ϵ, there are starting-points $P(t_n)$ from which $P(t)$ fails to make an ϵ-circuit of $\mathcal{S}$ in the arbitrarily large time T_n. A limit point Q of $\{P(t_n + \frac{1}{2}T_m)\}$ for $T_n \to \infty$ gives, by use of the Auxiliary Theorem and § 3 (v), a solution (Q) for which neither $L_\omega(Q)$ nor $L_\alpha(Q)$ coincides with $\mathcal{S}$. Since, as in (i), $L_\omega(Q)$ and $L_\alpha(Q)$ belong to $\mathcal{S}$, we have $L_\omega(Q) \subset \mathcal{S}$, $L_\alpha(Q) \subset \mathcal{S}$.

THEOREM IV. *A uniformly semi-pervasive solution is uniformly pervasive.*

Suppose that for a positively semi-pervasive solution (P), $P(t)$ passes within an arbitrary distance ϵ of each point of $L_\omega(P)$ for $t_0 \leq t \leq t_0 + T(\epsilon)$, where $T(\epsilon)$ is finite and independent of t_0. Let Q be any point of $L_\omega(P)$. Choose any positive ϵ and a sequence $\{t_i\}$ diminishing to $-\infty$. Then for each i there is a t'_i between t_i and $t_i + T(\epsilon)$ for which $P(t'_i)$ is within a distance ϵ of Q. The sequence $\{P(t'_i)\}$ has a limit point R in the closed ϵ-neighbourhood of Q; and R belongs to $L_\alpha(P)$ since $t'_i \to -\infty$ as $i \to \infty$. Since ϵ is arbitrary and $L_\alpha(P)$ is closed, Q belongs to $L_\alpha(P)$. Hence $L_\omega(P) \subseteqq L_\alpha(P)$.

But since (P) is positively semi-pervasive, $L_\alpha(P) \subseteqq L_\omega(P)$, (Theorem II). Hence $L_\alpha(P) = L_\omega(P)$ and (P) is pervasive; and our initial assumption shews it to be uniformly pervasive.

6. A property of uniformly pervasive (recurrent) trajectories.

THEOREM V. *If $P(t)$, $P(t+T)$ are any two current points on the same uniformly pervasive trajectory, the distance $\overline{P(t)P(t+T)}$ has, for T fixed and t variable, a positive lower bound.*

If the trajectory (P) is periodic the result is trivial. In the contrary case let P_0, P_1 be the fixed non-coincident points $P(0)$, $P(T)$; then a domain $\mathcal{D}_0$ surrounding P_0 has as its T-transform a domain $\mathcal{D}_1$ surrounding P_1. If $\mathcal{D}_0$ is infinitesimal so is $\mathcal{D}_1$, so we can and shall suppose that the distance between points of $\mathcal{D}_0$, $\mathcal{D}_1$ has a positive lower bound δ.

Since (P) is pervasive, $P(t)$ is within $\mathcal{D}_0$ at a sequence of instants

$$t = \cdots t_{-2}, t_{-1}, 0, t_1, t_2, \cdots, \tag{5}$$

where $t_i \to \infty$ and $t_{-i} \to -\infty$ as $i \to \infty$; and the points $P(t_i + T)$, being the T-transforms of $P(t_i)$, must be within $\mathcal{D}_1$.

Now given any two distinct current points $Q(t)$, $R(t)$, not necessarily on the same trajectory, there is a *positive* lower bound for the distance $\overline{Q(t)R(t)}$ during a given finite time interval $t' \leq t \leq t' + T'$, and for continuous variation of the starting points $Q(t')$, $R(t')$ this bound varies continuously. Hence if $Q(t')$, $R(t')$ be any points of $\mathfrak{D}_0$, $\mathfrak{D}_1$ respectively, there is a positive lower bound g for $\overline{Q(t)R(t)}$ for $t' \leq t \leq t' + T'$.

Let ϵ be the upper bound of the distances of points of $\mathfrak{D}_0$ from P_0 and, introducing the hypothesis of *uniform* pervasiveness, let T_ϵ be the time in which $P(t)$ uniformly performs an ϵ-circuit of the associated stratum; and let τ be the upper bound to the time any current point remains continuously within $\mathfrak{D}_0$ (so that when ϵ is small, τ is small and T_ϵ large). Then in the sequence (5) we can, for every i, take $\tau < t_{i+1} - t_i \leq T_\epsilon + \tau$. Choose $T' = T_\epsilon + \tau$, $Q(t') = P(t_i)$ and $R(t') = P(t_i + T)$, and we have

$$\overline{P(t)P(t+T)} \geq g \quad \text{for} \quad t_i \leq t \leq t_{i+1}.$$

Since this holds for all integers i, the lower bound g is valid for all t.

Q. E. D.

For recurrent trajectories of elementary continuous type, specified by periodic functions of arguments $\lambda_1 t + a_1$, $\lambda_2 t + a_2, \cdots$ ($\lambda_1, \lambda_2 \cdots$ constants mutually incommensurable; $a_1, a_2 \cdots$ arbitrary constants), there are the further properties:

(i) The ratio of the upper to the lower bound of $\overline{P(t)P(t+T)}$ has, for all T, a finite upper bound; so that if by taking T suitably large we have $P(T)$ close to $P(0)$, the points $P(t+T)$, $P(t)$ remain permanently close.

(ii) The same is true of the bounds of $\overline{P(t)Q(t)}$, where $Q(t)$ is in the stratum associated with (P) (which is a minimal set) but not on the trajectory (P).

For recurrent trajectories in general it seems that these statements are false, and that it is not even true that $\overline{P(t)Q(t)}$ has a lower bound distinct from zero.

7. Solutions associated with a non-uniformly pervasive or pervasive-asymptotic solution. With any pervasive or semi-pervasive solution (P) is associated its stratum $\mathcal{S}$, ($\mathcal{S} = L_\omega(P)$ or $L_\alpha(P)$), which is a perfect set of points in which the points of (P) are everywhere dense, and which contains all points of any solution having one point therein. If Q is any point of $\mathcal{S}$ we have therefore

$$L_\omega(Q) \subseteqq \mathcal{S}, \qquad L_\alpha(Q) \subseteqq \mathcal{S}.$$

We call Q

a *π-point* if $L_\omega(Q) = L_a(Q) = \mathcal{S}$,
a *σ-point* if $L_\omega(Q) \subset \mathcal{S}$ and $L_a(Q) \subset \mathcal{S}$,
a *μ-point* if either $L_\omega(Q) = \mathcal{S}$ and $L_a(Q) \subset \mathcal{S}$, or $L_\omega(Q) \subset \mathcal{S}$ and $L_a(Q) = \mathcal{S}$.

Every point of $\mathcal{S}$ belongs to one of these three classes; and all points of the same trajectory are of the same class. If (Q) consists of σ-points, $L_\omega(Q)$ and $L_a(Q)$ consist of σ-points. If (Q) consists of μ-points, that one of $L_\omega(Q)$, $L_a(Q)$ which is a part only of $\mathcal{S}$ consists of σ-points. If $\mathcal{S}$ contains one μ-point, the μ-points are everywhere dense therein.

If (P) is uniformly pervasive, $\mathcal{S}$ consists entirely (Theorem III (i)) of π-points.

If (P) is non-uniformly pervasive, $\mathcal{S}$ contains σ-points (Theorem III (ii)) as well as the π-points of (P). If (P) is pervasive-asymptotic, $\mathcal{S}$ contains μ-points (those of (P)), and, as just pointed out, σ-points. In these cases the uniformly pervasive solution or solutions which $\mathcal{S}$ must contain (Birkhoff [1], Theorem II) must consist of σ-points.

THEOREM VI. *The stratum associated with a non-periodic pervasive or pervasive-asymptotic solution contains an aggregate, having the cardinal number of the continuum, of pervasive trajectories which are everywhere dense therein, i. e. which consist of π-points.*

(i) Let (P) be a solution which is positively semi-pervasive, it being irrelevant whether negatively it is semi-pervasive or semi-asymptotic provided only it is not periodic. Then $L_\omega(P)$ is the associated stratum. Choose any point $P(t_0)$ of (P), this not being an equilibrium point since then (P) would be (degenerately) periodic. We can thus choose a neighbourhood $\mathfrak{N}$ of $P(t_0)$ and a positive constant τ such that any current point inside $\mathfrak{N}$ at $t = 0$ is outside $\mathfrak{N}$ at $t = \pm\tau$. Thus for any trajectory which traverses $\mathfrak{N}$, the portion within $\mathfrak{N}$ consists of a set of distinct arcs which is finite or enumerably infinite. In particular the trajectory (P) has infinitely many arcs within $\mathfrak{N}$; for since $P(t_0)$ belongs to $L_\omega(P)$, $P(t)$ must traverse $\mathfrak{N}$ infinitely often, and no two traverses are made on the same arc since (P) is not periodic.

(ii) Let us say that two closed neighbourhoods within $\mathfrak{N}$ are *not coupled* if there is no trajectory arc within $\mathfrak{N}$ having points common with both.

Let $\mathfrak{N}'$ be a neighbourhood of a point of (P), contained by $\mathfrak{N}$. Then $\mathfrak{N}'$ contains two closed neighbourhoods, centred at points of (P), which are not coupled. For since $\mathfrak{N}'$ is centred at a point of (P), which belongs to $L_\omega(P)$, $P(t)$ traverses $\mathfrak{N}'$ infinitely often on different arcs within $\mathfrak{N}$; the

closed ϵ-neighbourhoods of points P_0, P_1 of two such arcs, within $\mathfrak{N}'$, are then not coupled if ϵ is sufficiently small.

(iii) *Let there be assigned any closed neighbourhood $\bar{\mathfrak{N}}$ centred at any point of (P), and any positive numbers η, T; then there exists a neighbourhood $\phi(\bar{\mathfrak{N}}, \eta, T)$ centred at a point of (P), contained by $\bar{\mathfrak{N}}$, and such that any current point which is within $\phi(\bar{\mathfrak{N}}, \eta, T)$ at $t = 0$ is within the η-neighbourhood of $P(t_0)$ at two instants at least, one prior to $t = -T$ and the other subsequent to $t = T$.*[5] Let $P(t_1)$ be the centre and ϵ the radius of $\bar{\mathfrak{N}}$, i. e. $\bar{\mathfrak{N}}$ is the closed ϵ-neighbourhood of $P(t_1)$; and let $\epsilon' = \frac{1}{2}\epsilon$, $\eta' = \frac{1}{2}\eta$. Since $P(t_0)$ and $P(t_1)$ belong to $L_\omega(P)$ there exists a finite T' greater than T such that $P(t_0 + T')$ is within the ϵ'-neighbourhood of $P(t_1)$, and a finite T'' greater than T such that $P(t_0 + T' + T'')$ is within the η'-neighbourhood of $P(t_0)$. Thus a new current point $P'(t)$ on (P), which is coincident always with $P(t + t_0 + T')$, is within the ϵ'-neighbourhood of $P(t_1)$ at $t = 0$ and within the η'-neighbourhood of $P(t_0)$ at $t = -T'$ and $t = T''$.

Choose κ such that the distance $\overline{R(t)P'(t)}$ does not exceed η' for $-T' \leq t \leq T''$ provided only that $\overline{R(0)\ P'(0)} \leq \kappa$; and choose a positive κ' less than both κ and ϵ'. Then since $2\eta' = \eta$ any current point within the κ'-neighbourhood of $P'(0)$ i. e. of $P(t_0 + T')$ at $t = 0$ is within the η-neighbourhood of $P(t_0)$ at $t = -T', T''$. The former neighbourhood is contained by $\bar{\mathfrak{N}}$ since its radius and the distance of its centre from $P(t_1)$ are both less than ϵ', $= \frac{1}{2}\epsilon$; and it satisfies all the conditions required for $\phi(\bar{\mathfrak{N}}, \eta, T)$.

(iv) Proceeding now to the main theorem, let $\{\eta^{(i)}\}$ be a sequence of positive numbers diminishing to the limit 0, and $\{T^{(i)}\}$ an unbounded increasing sequence of positive numbers. Starting with the neighbourhood $\mathfrak{N}$ of (i), and the numbers $\eta^{(1)}, T^{(1)}$, select any closed neighbourhood $\bar{\mathfrak{N}}$ of $P(t_0)$ contained by $\mathfrak{N}$, and let a $\phi(\bar{\mathfrak{N}}, \eta^{(1)}, T^{(1)})$ be determined as in (iii). Then as in (ii) select two closed neighbourhoods $\bar{\mathfrak{N}}_0$, $\bar{\mathfrak{N}}_1$ centred at points of (P), contained by $\phi(\bar{\mathfrak{N}}, \eta^{(1)}, T^{(1)})$ and hence by $\bar{\mathfrak{N}}$, and not coupled. The radii of these may be chosen less than $\eta^{(1)}$. Then by (iii) any current point in $\bar{\mathfrak{N}}_0$ or $\bar{\mathfrak{N}}_1$ at $t = 0$ is within the $\eta^{(1)}$-neighbourhood of $P(t_0)$ at two instants at least, one prior to $t = -T^{(1)}$ and the other subsequent to $t = T^{(1)}$.

Next let there be determined a $\phi(\bar{\mathfrak{N}}_0, \eta^{(2)}, T^{(2)})$, and closed neighbourhoods $\bar{\mathfrak{N}}_{00}$, $\bar{\mathfrak{N}}_{01}$ of two points of (P), contained thereby, of radius less than $\eta^{(2)}$, and not coupled. Then $\bar{\mathfrak{N}}_{00}$, $\bar{\mathfrak{N}}_{01}$ are contained by $\bar{\mathfrak{N}}_0$, and any current point within either at $t = 0$ is within the $\eta^{(2)}$-neighbourhood of $P(t_0)$ at two instants at least, one prior to $t = -T^{(2)}$ and the other subsequent to $t = T^{(2)}$.

[5] This lemma is based on Birkhoff, *Dynamical Systems,* Chap. VII, sec. 11.

Next starting from $\bar{\mathfrak{N}}_1$ determine closed neighbourhoods $\bar{\mathfrak{N}}_{10}$, $\bar{\mathfrak{N}}_{11}$ therein with similar properties.

Proceeding thus, we have at the n-th stage 2^n closed neighbourhoods symbolized $\bar{\mathfrak{N}}_{i_1 i_2 \ldots i_n}$, where each i is either 0 or 1. Each of these is contained in all those of preceding stages whose symbols are obtained by curtailing the suffix $i_1 i_2 \cdots i_n$, is centred at a point of (P), and is of radius less than $\eta^{(n)}$; and a current point in any one of them at $t = 0$ is within the $\eta^{(n)}$-neighbourhood of $P(t_0)$ at instants prior to $t = -T^{(n)}$ and subsequent to $t = T^{(n)}$. No two of the neighbourhoods from any stages are coupled, since this would imply the coupling of corresponding neighbourhoods with curtailed symbols, which with sufficient curtailment is contrary to the construction.

(v) Corresponding to every infinite sequence $i_1 i_2, \cdots$, where each i is 0 or 1, we have a sequence of closed neighbourhoods

$$\bar{\mathfrak{N}}, \bar{\mathfrak{N}}_{i_1}, \bar{\mathfrak{N}}_{i_1 i_2}, \bar{\mathfrak{N}}_{i_1 i_2 i_3}, \cdots$$

each contained in its predecessor. There are $2^{\aleph_0}$ distinct sequences of this sort. Since the neighbourhoods are closed, each sequence determines a point $Q_{i_1 i_2 \ldots}$ common to all its members, there being only one such point since the radius $\eta^{(n)}$ tends to zero as $n \to \infty$. Each point Q belongs to $L_\omega(P)$ since it is a limit-point of the points of (P) at which the neighbourhoods of the corresponding sequence are centred. A current point which is at Q at $t = 0$ is for each n in the $\eta^{(n)}$-neighbourhood of $P(t_0)$ at instants prior to $t = -T^{(n)}$ and subsequent to $t = T^{(n)}$; hence $P(t_0)$ is both an α- and ω-limit point of (Q). Hence

$$(P) \subseteqq L_\alpha(Q), \qquad (P) \subseteqq L_\omega(Q);$$

and since the sets on the right are closed,

$$L_\omega(P) \subseteqq L_\alpha(Q), \qquad L_\omega(P) \subseteqq L_\omega(Q).$$

But since $Q \subset L_\omega(P)$,

$$L_\alpha(Q) \subseteqq L_\omega(P), \qquad L_\omega(Q) \subseteqq L_\omega(P).$$

Hence $L_\alpha(Q) = L_\omega(Q) = L_\omega(P)$, and (Q) consists of π-points.

Of the $2^{\aleph_0}$ points $Q_{i_1 i_2 \ldots}$, no two can be on the same trajectory arc within $\mathfrak{N}$, for this would imply the coupling of neighbourhoods of the sequences containing the two points. Hence there are at most $\aleph_0$ of the points on any one trajectory, and the aggregate of distinct trajectories consisting of π-points has the cardinal number of the continuum.

Q. E. D.

In virtue of this theorem the stratum associated with a pervasive-asymptotic solution can be regarded as associated with a solution which is pervasive, and

necessarily *non-uniformly* pervasive (Theorem III) since the stratum contains μ- and σ-points.

THEOREM VII. (i) *An arbitrary neighbourhood of any point of the stratum* $\mathcal{S}$ *associated with a non-uniformly pervasive or pervasive-asymptotic solution contains either* σ- *or* μ-*points of* $\mathcal{S}$.

(ii) *If the* σ-*points are not everywhere dense in* $\mathcal{S}$ *they are nowhere dense, and* $\mathcal{S}$ *contains a* μ-*trajectory of each species: that is, a trajectory* (R) *for which* $L_\omega(R) = \mathcal{S}$, $L_\alpha(R) \subset \mathcal{S}$, *and a trajectory* (R') *for which* $L_\alpha(R') = \mathcal{S}$, $L_\omega(R') \subset \mathcal{S}$.

(i) Let $\mathfrak{N}$ be a given neighbourhood of a point of $\mathcal{S}$. Let S be a σ-point of $\mathcal{S}$, and $P(t)$ a current point on a trajectory of $\mathcal{S}$ which consists of π-points. Then we can choose t_0 so that $P(t_0)$ is within $\mathfrak{N}$. $P(t_0)$ does not belong to $(S) + L_\omega(S) + L_\alpha(S)$ since every point of this set is a σ-point. This set being closed, $P(t_0)$ has hence a positive distance from it; and there is a neighbourhood $\mathfrak{N}'$ of $P(t_0)$ contained by $\mathfrak{N}$, the distances of points of which from points of (S) have a positive lower bound.

Let $\{P(t_i)\}$ be a sequence of points of (P) converging to the point S, all external to $\mathfrak{N}'$. $P(t)$ is outside $\mathfrak{N}'$ at $t = t_i$, and is inside at some subsequent instant since $P(t_0)$ belongs to $L_\omega(P)$. Hence there is a unique t'_i corresponding to and greater than each t_i, such that $P(t)$ is outside $\mathfrak{N}'$ for $t_i \leq t < t'_i$ and is on its boundary for $t = t'_i$.

Let R be a limit point of the sequence $\{P(t'_i)\}$; then $R(t)$, the current point which is at R at $t = 0$, is for any t a limit point of the sequence $\{P(t'_i + t)\}$. Now since the distance between any points of (S) and of $\mathfrak{N}'$ has a positive lower bound, it follows from the Auxiliary Theorem that $| t'_i - t_i | \to \infty$ as the distance $\overline{SP(t_i)} \to 0$, and hence as $i \to \infty$; and with $t'_i > t_i$ this implies $t'_i - t_i \to \infty$. Hence for any fixed negative t we have, for all sufficiently large i, $t_i < t'_i + t < t'_i$; so that $P(t'_i + t)$ is outside $\mathfrak{N}'$, and hence $R(t)$ not within $\mathfrak{N}'$. Thus $L_\alpha(R)$ has no point within $\mathfrak{N}'$, and (R) is a trajectory consisting of σ- or μ-points. The point R lies on the boundary of $\mathfrak{N}'$ since it is a limit point of the $P(t'_i)$ which lie there, and is the required σ- or μ-point within $\mathfrak{N}$.

Note: It is not proved that S belongs to $L_\alpha(R)$, nor that $R(t)$ is for any positive t within $\mathfrak{N}'$.

(ii) Suppose now that the σ-points are not everywhere dense in $\mathcal{S}$, and that $\mathfrak{N}$ is a neighbourhood of a point of $\mathcal{S}$ free from σ-points. Let $\mathfrak{D}$ be any neighbourhood of any point of $\mathcal{S}$.

We can choose t_0, t_1 so that $P(t_0)$, $P(t_1)$ lie within $\mathfrak{N}$, $\mathfrak{D}$ respectively.

A sufficiently small neighbourhood $\mathfrak{D}'$ of $P(t_1)$ has its $(t_0 - t_1)$-transform contained by $\mathfrak{N}$, and $\mathfrak{D}'$ contains no σ-points since any transform of a σ-point is a σ-point. Hence there is a sub-neighbourhood of $\mathfrak{D}$, centred at a point of $\mathcal{S}$, free of σ-points; and the σ-points are nowhere dense in $\mathcal{S}$.

It follows that the point R defined in (i), being within $\mathfrak{N}$, must be a μ-point. Hence (R) consists of μ-points, and since $L_\alpha(R) \subset \mathcal{S}$ we must have $L_\omega(R) = \mathcal{S}$. In a similar manner, by taking t'_i less than t_i, we find a trajectory (R') with $L_\alpha(R') = \mathcal{S}$, $L_\omega(R') \subset \mathcal{S}$.

Q. E. D.

In view of this theorem there are three logically possible types of strata $\mathcal{S}$ associated with non-uniformly pervasive solutions, the characteristic properties being

(a) the σ-points are nowhere dense and the μ-points everywhere dense in $\mathcal{S}$,

(b) both σ- and μ-points are everywhere dense in $\mathcal{S}$,

(c) the σ-points are everywhere dense and there are no μ-points in $\mathcal{S}$.

If the solution with which $\mathcal{S}$ was originally associated is pervasive-asymptotic, $\mathcal{S}$ can be of type (a) or (b) only. The type (b) can be subdivided according as there are μ-trajectories of both species or of one species only.

In §§ 9-11 below, all these types are illustrated, though only in the cases of type (a) are the X_i in the differential equations regular-analytic. The type (b) is illustrated by the quasi-ergodic systems of geodesics on closed surfaces of negative curvature which have been given by Artin [1] and others; in such cases the stratum is not of course the surface on which the geodesics lie, but a related three-dimensional manifold.

(iii) In the case where the σ-points are everywhere dense in $\mathcal{S}$, let us call a point Q of $\mathcal{S}$ a *σ-point of class η* if η is the upper bound of distances of points of $\mathcal{S}$ from the trajectory (Q); and let Σ_η be the set of all σ-points of $\mathcal{S}$ whose class is not less than η. Then it is easy to establish the following facts:

(a) *If Q belongs to Σ_η then every point of (Q), $L_\omega(Q)$ and $L_\alpha(Q)$ belongs to Σ_η.*

(b) *There is no point of Σ_η in a sufficiently small neighbourhood of any π-point or μ-point or σ-point of class less than η.*

(c) *The set Σ_η is closed.*

(d) *Either an arbitrary neighbourhood of Σ_η contains complete σ-trajectories (whose class-number is less than but arbitrarily near η), or there are trajectories of positively (negatively) semi-asymptotic type whose ω- (α-) limit set belongs to Σ_η; in the latter alternative a set of suitably selected points, one on each*

of the semi-asymptotic trajectories, is closed, but the set of all points on these trajectories is not necessarily closed.

Birkhoff has called the system of trajectories filling a closed connected region $\mathcal{S}$ *transitive* if arbitrary neighbourhoods of any two points of the region are traversed by the same trajectory of the system; and he has proved (*Dynamical Systems,* Chap. VII, Sec. 11) that in this case $\mathcal{S}$ contains trajectories whose ω- and α-limit sets coincide with $\mathcal{S}$. Thus $\mathcal{S}$ is a stratum, and the theories of the structure of strata and of transitive systems are identical. In Birkhoff's terminology, the trajectories which consist of π-points are *general,* and those which consist of σ- or μ-points are *special.* For the non-trivial case in which the system contains special trajectories, it may be observed that Theorem VI contains a new result concerning the power of the aggregate of general trajectories, and that Theorem VII reproduces Birkhoff's theorems on the special trajectories (*Dynamical Systems,* Chap. VII, Sec. 9) with a somewhat more direct proof.

The analysis of $\mathcal{S}$ into π-, μ- and σ-points may be of some significance in connection with the problem of metrical transitivity. If $\mathcal{S}$ is a regular hypersurface on which measure is suitably defined and on which there is a positive integral-invariant, it is known (Carathéodory [1]) that the μ-points are of measure 0. The question then arises whether the σ-points are of measure 0, which is of course necessary for the metrical transitivity of $\mathcal{S}$. I have not been able to do more than reduce this question, by means of (c) in (iii) above, to the following: Is a *closed* invariant set which is a part only of $\mathcal{S}$ necessarily of measure 0?

Construction of Systems of Differential Equations Having Non-Uniformly Pervasive and Pervasive-Asymptotic Solutions.

8. As starting-point of the construction we take a system of differential equations in two dependent variables r, z:

$$dr/ds = R(r, z), \qquad dz/ds = Z(r, z) \tag{6}$$

which possesses a (real) periodic solution

$$r = f(s), \qquad z = g(s). \tag{7}$$

Such systems are easy to construct, by starting with an equation

$$F(r, z) = \text{constant} \tag{8}$$

which, with r and z as rectangular coördinates, represents a family of closed

curves without singular points, and eliminating the constant by differentiation. For example, from

$$r^2 + z^2 = \text{constant} \tag{9}$$

we may pass to the differential equations

$$dr/ds = -z, \qquad dz/ds = r$$

whose solutions

$$r = c \cos s, \qquad z = c \sin s$$

are periodic, and reproduce the circles (9).

The properties postulated for the functions occurring in (6) and (7) are as follows:

(i) $R(r, z)$ and $Z(r, z)$ are analytic functions of r, z, real and regular at all points $(f(s), g(s))$ for which s is real.

(ii) $f(s)$ and $g(s)$ are analytic functions of s, real and regular for all real s, and periodic with period S;

(iii) the lower bounds of $f(s)$ and $g(s)$, for s real, are positive. The properties (i), (ii) are here stated in a form convenient for subsequent requirements, and without regard for the fact that they have some logical connection. From them, together with the general theorem of uniqueness of solution of a system such as (6), there follows:

(iv) the closed curve specified by (7) has no singular points.

The property (iii) is, in geometrical terms, that the curve (7) lies entirely in the positive quadrant of the rz-plane.

We now pass to a system with three dependent variables x, y, z by adjoining to (6) a differential equation

$$d\theta/ds = 2\pi/T$$

where T is any constant such that S/T is an irrational number, and putting

$$x = r \cos \theta, \qquad y = r \sin \theta. \tag{10}$$

In the space in which x, y, z are rectangular coördinates, r, θ, z are cylindrical coördinates, and the equations (7) represent a torus $\mathcal{S}$ obtained by rigid rotation of the curve (7) about the axis $r = 0$; and in virtue of (iii), (iv) above, $\mathcal{S}$ has no singular points.

In terms of x, y, z the system is

$$dx/ds = xR/r - 2\pi y/T, \quad dy/ds = yR/r + 2\pi x/T, \quad dz/ds = Z, \tag{11}$$

where we are to put $r = +\sqrt{(x^2 + y^2)}$; its right members are (from (i) and

(iii)) single-valued regular functions of x, y, z for (x, y, z) on $\mathcal{S}$. $\mathcal{S}$ is made up of a family of the trajectories of (11), which may be specified by the corresponding solutions

(12) $x = f(s)\cos(2\pi s/T + \theta_0)$, $y = f(s)\sin(2\pi s/T + \theta_0)$, $z = g(s)$,

θ_0 being an arbitrary constant.

It is well known that, since S/T is irrational, this family is a minimal set of recurrent, or uniformly pervasive, trajectories each of which has $\mathcal{S}$ for associated stratum. Each trajectory, if followed in either sense, spirals indefinitely round $\mathcal{S}$ without singularity, and has infinitely many arcs lying in an arbitrary neighbourhood of any point of $\mathcal{S}$. Each trajectory may be specified by any value of the constant θ_0 chosen from an enumerably infinite set; and there are an infinity of distinct trajectories, of the cardinal number of the continuum, one of which passes through any assigned point of $\mathcal{S}$.

9. Choose any point A of the surface $\mathcal{S}$, say (x_1, y_1, z_1). Let $E(x, y, z)$ be an analytic function, real and regular for all (x, y, z) on $\mathcal{S}$ and positive at all these points except A, where it vanishes; for example

$$E \equiv (x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2.$$

We shall shew that *the differential equations*

(13) $dx/dt = E(xR/r - 2\pi y/T)$, $dy/dt = E(yR/r + 2\pi x/T)$, $dz/dt = EZ$

have certain solutions, viz. those which lie in $\mathcal{S}$, which are either non-uniformly pervasive or pervasive-asymptotic.

Since (13) is related to (11) by the substitution

$$dt = ds/E,$$

the trajectories of (13) are geometrically coincident with those of (11); and for the solutions carried by these trajectories we have to determine the manner in which t is correlated with their points. For the solutions in $\mathcal{S}$ we have

$$(14) \qquad t = \int_{s_0}^{s} \frac{ds}{E\{f(s)\cos(2\pi s/T + \theta_0), f(s)\sin(2\pi s/T + \theta_0), g(s)\}},$$

where s_0 is an arbitrary constant effectively additive with t.

First let θ_0 have such a value that the curve (12) passes through A, the value, necessarily unique, of the parameter s at this point being s_1 say. Then for $s = s_1$ the denominator E on the right of (14) vanishes, while for $s \neq s_1$ it is positive with a positive upper bound. Since E is a regular function of x, y, z which are, by (12), regular functions of s, E is a regular function of s; and near s_1 we have a convergent expansion

$$E = a(s - s_1)^2 + b(s - s_1)^3 + \cdots$$

where, because E is never negative on $\mathcal{S}$, the leading term has a positive coefficient and an even exponent. Hence the integral in (14) diverges at $s = s_1$ like an odd negative-integral power of $(s_1 - s)$.

Thus if $s_0 > s_1$, (14) gives t as a single-valued strictly increasing function of s defined for $s > s_1$, and $t \to -\infty$ as $s \to s_1 + 0$; while since $1/E$ has a positive lower bound, $t \to \infty$ as $s \to \infty$. If $s_0 < s_1$, t is a single-valued strictly increasing function of s defined for $s < s_1$, and $t \to \infty$ as $s \to s_1 - 0$, $t \to -\infty$ as $s \to -\infty$. In both cases, inversion gives s as a single-valued strictly increasing function of t defined for all real t; if $s_0 > s_1$ we have $s \to s_1$ as $t \to -\infty$, and $s \to \infty$ as $t \to \infty$; if $s_0 < s_1$ we have $s \to -\infty$ as $t \to -\infty$, and $s \to s_1$ as $t \to \infty$. Hence the curve (12) splits into three complete trajectories of (13), for which respectively $s = s_1$ (giving the equilibrium point A), $s > s_1$, and $s < s_1$. Any solution carried by the portion for which $s > s_1$ has the point A as its α-limit set and, since $s \to \infty$ as $t \to \infty$, the whole surface $\mathcal{S}$ as its ω-limit set. For the other portion the ω-limit set is A and the α-limit set is $\mathcal{S}$. Thus for (13) we have just one pervasive-asymptotic trajectory of each species, each of which carries of course the infinitely many trivially different solutions corresponding to different values of s_0.

Now let θ_0 have such a value that the curve (12) passes arbitrarily near to but not through A. Then in (14), E vanishes for no value of s though its lower bound is 0, and t is a single-valued strictly increasing function of s defined for all real s; and since $1/E$ has a positive lower bound, $t \to \pm\infty$ as $s \to \pm\infty$. Inversion gives s as a single-valued strictly increasing function of t, defined for all real t, with $s \to \pm\infty$ as $t \to \pm\infty$. Thus the curve furnishes one trajectory of (13), any solution carried by which has the whole surface $\mathcal{S}$ as its α- and as its ω-limit set. The solution (or trajectory) is hence pervasive, with $\mathcal{S}$ as its stratum. That it is non-uniformly pervasive follows most easily from Theorem III (i) since we already know $\mathcal{S}$ to contain σ- and μ-points. The mechanism of the non-uniformity is of course that in (14) the denominator E is arbitrarily small for the infinitely many values of s corresponding to sufficiently near approaches to A; and it may be proved directly that if a current point enters the ϵ-neighbourhood of A, where $\epsilon \sim 0$, it remains within a fixed neighbourhood of A for a time of order $1/\epsilon$.

We see therefore that $\mathcal{S}$ is the stratum associated with each of the non-uniformly *pervasive* or *pervasive*-asymptotic trajectories of (13) contained thereby. It contains one σ-point A, two trajectories made up of μ-points, and an infinity, of the cardinal number of the continuum, of trajectories made up of π-points.

10. The example of § 9 is in accord with the definitions of §§ 3-5, but is, so to speak, psychologically weak: the equilibrium point A is removable by change of the independent variable, and it might be held that the definitions should be modified so as to presuppose such removal. It is however easy to pass to an example free from this possible objection.

Adjoin to (13) the differential equation

$$d\psi/dt = 2\pi/U,$$

where U is a positive constant, and let new variables u, v be introduced by the relations

$$u = z\cos\psi, \quad v = z\sin\psi, \quad z = +\sqrt{(u^2+v^2)}. \tag{15}$$

We obtain the system of the fourth order

$$\left.\begin{array}{ll} dx/dt = E(xR/r - 2\pi y/T), & dy/dt = E(yR/r + 2\pi x/T) \\ du/dt = uEZ/z - 2\pi v/U, & dv/dt = vEZ/z + 2\pi u/U \end{array}\right\}, \tag{16}$$

of which ∞^1 solutions are got by adjoining to any solution of (13) the equation

$$\psi = 2\pi t/U + \psi_0.$$

If the solution of (13) lies in $\mathcal{S}$, the solutions of (16) thus generated lie in the hypersurface $\mathcal{H}$ obtained by rotating $\mathcal{S}$ about the plane $u = v = 0$ (or $z = 0$) in the four dimensional space in which x, y, u, v, are rectangular coördinates. $\mathcal{H}$ is without singular points since by the initial assumption (iii) of § 8 $\mathcal{S}$ has no point common with the plane $z = 0$; and $\mathcal{S}$ is the section of $\mathcal{H}$ by the half-hyperplane $\psi = 0$. Any solution (Q) in $\mathcal{H}$ has in common with $\mathcal{S}$ the points

$$\cdots P(t_0 - 2U),\ P(t_0 - U),\ P(t_0),\ P(t_0 + U),\ P(t_0 + 2U), \cdots$$

occupied at equidistant time-intervals by some current point $P(t)$ which moves in $\mathcal{S}$ according to the differential equations (13). The points of $L_\omega(Q)$, $L_\alpha(Q)$ in $\mathcal{S}$ are the limit points of the respective sequences

$$P(t_0), P(t_0 + U), P(t_0 + 2U), \cdots, \tag{17}$$

$$P(t_0), P(t_0 - U), P(t_0 - 2U), \cdots, \tag{18}$$

and the complete sets $L_\omega(Q)$, $L_\alpha(Q)$ consist of the trajectories of (16) through these limit points. We call (17) and (18) the *positive* and *negative U-sets based on* $P(t_0)$ and desire therefore to know their derived sets when $P(t_0)$ is any assigned point of $\mathcal{S}$.

If $P(t_0)$ is at A all the points of (17), (18) are at A, and we have for the system (16) a periodic trajectory $\mathcal{T}$ carrying the ∞^1 trivially-different solutions

$$x = x_1,\ y = y_1,\ u = z_1\cos(2\pi t/U + \psi_0),\ v = z_1\sin(2\pi t/U + \psi_0). \tag{19}$$

If $P(t)$ asymptotes to A as $t \to \infty$, any associated positive U-set con-

verges to A, and the generated solutions of (16) have the periodic trajectory $\mathcal{J}$ for ω-limit set. There are ∞^2 such solutions, carried by ∞^1 trajectories. There are similarly ∞^1 trajectories of (16) having $\mathcal{J}$ as α-limit set.

The only other, and most general, case is that where $P(t)$ passes arbitrarily near to each point of $\mathcal{S}$ as $t \rightarrow \infty$ (or as $t \rightarrow -\infty$). In this case it is clear that *any associated positive (or negative) U-set cannot converge to A,* but it is not immediately evident whether or not such a set is everywhere dense in $\mathcal{S}$. [If for instance, altering our hypotheses, we take $E \equiv 1$ on $\mathcal{S}$, such a set is not or is everywhere dense in $\mathcal{S}$ according as U is or is not commensurably related to S and T.] It can be proved that *every U-set except those which converge to A is everywhere dense in $\mathcal{S}$, whatever the value of the constant U, provided that S/T is a transcendental number of a certain type and that the analytic function E of the complex variable s is suitably chosen.*[6] The proof of this is too long to be given here, and the following result is sufficient for our purpose:

THEOREM VIII. *For any choice, in accordance with the conditions previously stipulated, for the functions and constants occurring in the right members of the equations* (13), *there is a value of U arbitrarily near to any given positive number U_0 for which every U-set in $\mathcal{S}$, except those which converge to the point A, is everywhere dense in $\mathcal{S}$.*

Assuming this for the moment, it follows that any positively (negatively) semi-pervasive solution of (13) generates ∞^1 solutions of (16) which are positively (negatively) semi-pervasive with $\mathcal{H}$ as the associated stratum. Thus $\mathcal{H}$ is made up of the following sets: (a) the periodic trajectory $\mathcal{J}$; (b) ∞^1 trajectories (Q) for which $L_\omega(Q) = \mathcal{J}$, $L_\alpha(Q) = \mathcal{H}$, and ∞^1 trajectories (Q') for which $L_\alpha(Q') = \mathcal{J}$, $L_\omega(Q') = \mathcal{H}$;and (c) ∞^2 trajectories (Q'') for which $L_\omega(Q'') = L_\alpha(Q'') = \mathcal{H}$. It follows as for the system (13) that the trajectories (Q'') are non-uniformly pervasive.

Proof of Theorem VIII. (i) If $P(t)$ is any current point in $\mathcal{S}$, we call the aggregate of points $\{P(t)\}$ for which $t' \leq t \leq t' + \tau$ a *closed*[7] *trajectory-arc of duration* τ. If Q is any point of $\mathcal{S}$, other than A, let $\mathcal{C}$ be an analytic arc through Q lying in $\mathcal{S}$ and nowhere tangential to a trajectory; the U-transform of $\mathcal{C}$ is then an analytic arc $\mathcal{C}_U$, and the closed trajectory-arcs which stretch from $\mathcal{C}$ to $\mathcal{C}_U$ are all of duration U.

---

[6] If $S/T = \frac{1}{a_1 +}\frac{1}{a_2 +}\frac{1}{a_3 +}\cdots$ with $a_{n+1} > 2q_n/\epsilon_{n+1}$, where p_n/q_n is the n-th convergent and $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$, asymptotic formulae connecting the values of the integral in (14) over suitably chosen ranges can be developed by deformation of the path of integration, and lead to the stated result.

[7] i. e. "closed" in the sense that the end points belong to the arc.

If $P(t)$ passes arbitrarily near to each point of $\mathcal{S}$ as $t \to \infty$, the positive U-set based on any point $P(t_0)$ of its trajectory has a limit point on any closed trajectory-arc QQ_U of duration U. For the half-trajectory described by $P(t)$ for $t \geq t_0$ cuts $\mathcal{C}$, at $P(t')$ say, in an arbitrary neighbourhood of Q. $P(t' + U)$ then lies on $\mathcal{C}_U$, and if $P(t')$ is sufficiently near to Q the closed trajectory-arc from $P(t')$ to $P(t' + U)$ lies in an arbitrary neighbourhood of the arc QQ_U. The first named arc contains one point of the U-set

$$P(t_0), P(t_0 + U), P(t_0 + 2U), \cdots$$

which is in question, or two points if $t' \equiv t_0 \pmod{U}$. Hence in an arbitrary neighbourhood of the closed arc QQ_U there are infinitely many points of the U-set, which has therefore a limit point on the arc.

Similarly, *if $P(t)$ passes arbitrarily near to each point of $\mathcal{S}$ as $t \to -\infty$, the negative U-set based on any point $P(t_0)$ of its trajectory has a limit point on any closed trajectory-arc of duration U.*

(ii) *If $P(t)$ passes arbitrarily near to each point of $\mathcal{S}$ as $t \to \infty$, $P(t_0)$ is a given point of its trajectory, and U_0 is a given positive number, then we can choose U as near as we please to U_0 such that the positive U-set based on $P(t_0)$ has every point of the trajectory-arc QQ_U as a limit point.*

We shew how to choose U so that an arbitrary neighbourhood of any point of the trajectory-arc from Q to its $(2U_0)$-transform Q_{2U_0} contains infinitely many points of the positive U-set based on $P(t_0)$. For this purpose let $\mathcal{C}$ be an analytic arc through Q not tangential to any trajectory, ϕ the arc-length on $\mathcal{C}$ measured from Q, and $\mathcal{C}_r$ the r-transform of $\mathcal{C}$. Any trajectory-arc near QQ_{2U_0} can then be referred to the value of ϕ at the point at which it cuts $\mathcal{C}$. A set of domains $\mathcal{D}_n$ in $\mathcal{S}$ is then defined as follows, η being a positive constant:

$\mathcal{D}_1$ bounded by arcs of	$\mathcal{C}$,	$\mathcal{C}_{U_0}$	and of trajectories	$\phi = \eta, -\eta$
$\mathcal{D}_2$ " " " "	$\mathcal{C}_{U_0}$,	$\mathcal{C}_{2U_0}$	" " "	$\phi = \pm \frac{1}{2}\eta$
$\mathcal{D}_3$ " " " "	$\mathcal{C}$,	$\mathcal{C}_{1/2 U_0}$	" " "	$\phi = \pm \frac{1}{3}\eta$
$\mathcal{D}_4$ " " " "	$\mathcal{C}_{1/2 U_0}$,	$\mathcal{C}_{U_0}$	" " "	$\phi = \pm \frac{1}{4}\eta$
$\mathcal{D}_5$ " " " "	$\mathcal{C}_{U_0}$,	$\mathcal{C}_{3/2 U_0}$	" " "	$\phi = \pm \frac{1}{5}\eta$
$\mathcal{D}_6$ " " " "	$\mathcal{C}_{3/2 U_0}$,	$\mathcal{C}_{2U_0}$	" " "	$\phi = \pm \frac{1}{6}\eta$
$\mathcal{D}_7$ " " " "	$\mathcal{C}$,	$\mathcal{C}_{1/4 U_0}$	" " "	$\phi = \pm \frac{1}{7}\eta$.
. . . .	. . .	. . .	. . . .	. . .

An arbitrary neighbourhood of any point of the arc QQ_{2U_0} then contains infinitely many of the $\mathcal{D}_n$.

Let $2t_n$ be the time-interval between the end-arcs of $\mathcal{D}_n$, so that

$$t_1 = t_2 = \tfrac{1}{2}U_0,\ t_3 = \cdots = t_6 = \tfrac{1}{4}U_0,\ t_7 = \cdots = t_{14} = \tfrac{1}{8}U_0, \text{ etc.};$$

and

(20) $$t_{n+1} \leq t_n < 2U_0.$$

Call the arc in $\mathcal{D}_n$ which is at equal time-intervals from its end-arcs its *axis*; e. g. $\mathcal{C}_{1/2U_0}$ is the axis of $\mathcal{D}_1$.

By hypothesis $P(t)$ traverses each $\mathcal{D}_n$ for arbitrarily large values of t, so there are arbitrarily large integers i_n such that the point $P(t_0 + i_nU)$ of the U-set based on $P(t_0)$ lies on a trajectory-arc traversing $\mathcal{D}_n$ at a time-interval less than U from its axis. If U increases by δU, t_0 remaining fixed, $P(t_0 + i_nU)$ advances along the trajectory (P) through an arc of duration $i_n\delta U$; so we can choose δU, with $|\delta U| < U/i_n$, such that $P(t_0 + i_n(U + \delta U))$ lies in $\mathcal{D}_n$ on its axis. By giving U a succession of increments from the starting-value U_0 we thus move points of the U-set, of suitably chosen ranks $i_1, i_2, \cdots$, in succession on to the axes of $\mathcal{D}_1, \mathcal{D}_2, \cdots$. The alteration in U at the n-th stage produces a consequential displacement of each of the points $P(t_0 + i_1U), \cdots, P(t_0 + i_{n-1}U)$; but by taking the ratios i_2/i_1, $i_3/i_2, \cdots$ sufficiently large we secure that, for each j, $P(t_0 + i_jU)$ never gets moved out of $\mathcal{D}_j$.

A suitable rule for choosing $i_1, i_2, \cdots$ is as follows: Choose a constant κ between 0 and $\frac{1}{2}$. Choose i_1 greater than $2/\kappa$ and such that $P(t_0 + i_1U_0)$ lies on a trajectory-arc traversing $\mathcal{D}_1$, at a time-interval less than U_0 from its axis. Choose $U_1, = U_0 + \delta_0U$, such that $P(t_0 + i_1U_1)$ lies on the axis of $\mathcal{D}_1$. Then

$$|\delta_0U| < U_0/i_1 < 2U_0/i_1 < \kappa U_0 < U_0, \qquad 0 < U_1 < 2U_0.$$

Suppose that $i_1, \cdots, i_{n-1}$ have been chosen in succession such that

(a) $U_{n-1} = U_0 + \delta_0U + \delta_1U + \cdots + \delta_{n-2}U$ with
$$|\delta_jU| < \kappa^{j+1}U_0, \qquad (j = 0, 1, \cdots, n-2),$$

so that

(21) $$|U_{n-1} - U_0| < \kappa U_0/(1-\kappa) < U_0;$$

(22) (b) $i_j > 2U_0 i_{j-1}/\kappa t_{j-1}, \qquad (j = 2, 3, \cdots, n-1);$

(c) $P(t_0 + i_{n-1}U_{n-1})$ lies on the axis of $\mathcal{D}_{n-1}$;

(d) $P(t_0 + i_jU_{n-1})$, for $j = 1, 2, \cdots, n-2$, lies in $\mathcal{D}_j$ at a time-interval from its axis less in absolute value than

$$(\kappa + \kappa^2 + \cdots + \kappa^{n-1-j})t_j.$$

Choose i_n greater than $2U_0i_{n-1}/(\kappa t_{n-1})$ and such that $P(t_0 + i_nU_{n-1})$ lies on a trajectory-arc traversing $\mathcal{D}_n$, at a time-interval from its axis less than U_{n-1}. Choose $U_n, = U_{n-1} + \delta_{n-1}U$, such that $P(t_0 + i_nU_n)$ lies on the axis of $\mathcal{D}_n$. Then from (21)

$$|\delta_{n-1}U| < U_{n-1}/i_n < 2U_0/i_n;$$

and the time-interval between $P(t_0 + i_j U_{n-1})$ and $P(t_0 + i_j U_n)$ is $i_j \delta_{n-1} U$, with an absolute value less than $2U_0 i_j / i_n$. From (20), (22) we have, for $j = 2, 3, \cdots, (n-1)$,

$$i_{j-1}/i_j < \kappa(t_{j-1}/2U_0) < \kappa,$$

and the same relation holds for $j = n$ in virtue of the choice of i_n. Hence

$$2U_0/i_n = 2U_0/i_1 \cdot i_1/i_2 \cdots i_{n-1}/i_n < \kappa U_0 \cdot \kappa^{n-1} = \kappa^n U_0,$$

and for $j = 1, 2, \cdots, n-1$

$$\begin{aligned} i_j/i_n &= i_j/i_{j+1} \cdot i_{j+1}/i_{j+2} \cdots i_{n-1}/i_n \\ &< (\kappa t_j/2U_0) \cdot \kappa^{n-j-1} = \kappa^{n-j} t_j/2U_0. \end{aligned}$$

Thus $| \delta_{n-1} U | < \kappa^n U_0$, and the time-interval of $P(t_0 + i_j U_n)$ from the axis of $\mathfrak{D}_j$ is less in absolute value than $(\kappa + \cdots + \kappa^{n-j}) t_j$. The properties (a), (b), (c), (d) are thus by induction established for all n. The sequence $\{U_n\}$ converges to the desired value U, for which

$$| U - U_0 | < \kappa U_0/(1-\kappa) < U_0, \qquad 0 < U < 2U_0,$$

while the time-interval of $P(t_0 + i_j U)$ from the axis of $\mathfrak{D}_j$ is less in absolute value than $\kappa t_j/(1-\kappa)$, which is less than t_j, so that $P(t_0 + i_j U)$ lies in $\mathfrak{D}_j$. For a sufficiently small choice of κ, U is as near as we please to U_0.

It follows immediately that *every point of the trajectory-arc QQ_{2U_0} is a limit point of the U-set based on $P(t_0)$*; and there is the further relevant fact that *this arc contains a closed arc QQ_U of duration U.*

(iii) *With the value of U just chosen, every U-set is everywhere dense in $\mathcal{S}$, except the positive (negative) U-sets on the trajectory positively (negatively) asymptotic to A.*

If any U-set has a limit point R, the Auxiliary Theorem shews that the t-transform of the set has the t-transform of R as a limit point, for any t. In particular the U-transform and the $(-U)$-transform of R are limit points *of the original set.*

Since the U-set $P(t_0), P(t_0 + U), \cdots$ which is in question in (ii) above has every point of the trajectory-arc QQ_U as a limit point, and Q_U is the U-transform of Q, it follows that every point of (Q) is a limit point. But $Q(t)$ passes arbitrarily near to each point of $\mathcal{S}$ as $t \to \infty$, or $t \to -\infty$, or both, since every solution (except A) in $\mathcal{S}$ has $\mathcal{S}$ as its ω- or α-limit set, or both. Hence there is one U-set $P(t_0), P(t_0 + U), \cdots$ whose derived set is $\mathcal{S}$.

From what was first pointed out it follows that for any t' the positive U-set based on $P(t_0 + t')$ has for derived set the t'-transform of $\mathcal{S}$, which coincides with $\mathcal{S}$.

Now any U-set K which is not of the type excepted in the enunciation

has by (i) a limit point, $P(t_0 + t')$ say, on (P). Hence as before K has each of the points

$$P(t_0 + t'),\ P(t_0 + t' + U),\ P(t_0 + t' + 2U), \cdots$$

as a limit point. These points having just been seen to be everywhere dense in $\mathcal{S}$, K must be everywhere dense in $\mathcal{S}$.

Q. E. D.

It is of interest to note that the functions R, Z, E can be so chosen, in accordance with the conditions of §§ 8, 9, that the right members of the equations (13) are polynomials in x, y, z, and those of the equations (16) polynomials in x, y, u, v. The simplest choice of $F(r, z)$ appears to be

$$F(r, z) = r^2 + z^2 + 1/r^2 + 1/z^2.$$

If $c > 4$ the curve $F = c$ consists of ovals symmetrically placed one in each quadrant; for $c = 4$ we get the four acnodes $r = \pm 1$, $z = \pm 1$; and the axes $r = 0$, $z = 0$ correspond to $c = \infty$. Eliminating c, the system (6) is

$$dr/ds = r^3(z^4 - 1) = R, \qquad dz/ds = z^3(1 - r^4) = Z.$$

The solutions for $c > 4$ are periodic with a period S which is an analytic function of c, and any one of these solutions can serve as the basis (7) for the construction. The right members of (13), (16) will be polynomials as stated if E is a polynomial in x, y, z^2; for example

$$E = (x - x_1)^2 + (y - y_1)^2 + (z^2 - z_1^{\,2})^2.$$

11. The principles exemplified in §§ 9, 10 can be used to construct more complicated examples. The function E can be chosen, as a polynomial, to vanish at any finite number of assigned points $A_1, \cdots, A_n$ of $\mathcal{S}$ and be positive elsewhere on $\mathcal{S}$, and the points $A_1, \cdots, A_n$ may be on the same or on different trajectories of the system (11). If $A_1, \cdots, A_m$ are on the same trajectory, this splits into $(2m + 1)$ complete trajectories of the system (13), namely the equilibrium points $A_1, \cdots, A_m$, the finite open arcs $A_1A_2, \cdots, A_{m-1}A_m$, and two semi-infinite arcs. The points of the latter become μ-points, and $A_1, \cdots, A_m$ and the points of the finite arcs become σ-points.

If the points A are infinite in number but not everywhere dense in $\mathcal{S}$, an E which is not identically zero must cease to be *regular-analytic* at their limit points, but can be chosen so as to be *regular*, in the sense defined in § 1, at all points. With this sort of E we can exemplify the types (b), (c) of § 7; the examples already given are of type (a), with the σ-points nowhere dense in $\mathcal{S}$.

Taking the trajectory (12) of the system (11) for which θ_0 has an assigned value, let equilibrium points $A_1, A_2, \cdots$ be introduced at the points

$$s = s_1, s_2, \cdots, (s_1 < s_2 < s_3 \cdots),$$

where s_n is so chosen that the sequence $\{A_n\}$ has A_1 as its sole limit point; this implies that $s_n \to \infty$ as $n \to \infty$. E is of course to vanish at $A_1, A_2, \cdots$ and be positive elsewhere on $\mathcal{S}$. Then the half-trajectory $s \geq s_1$ of (11) furnishes an infinite number of complete trajectories of (13), viz. the points A_n and the finite open arcs $A_n A_{n+1}$, and each of these consists of σ-points. The half-trajectory $s < s_1$ consists as before of μ-points. Hence the stratum $\mathcal{S}$ is of type (b), with both σ- and μ-points everywhere dense therein, and all μ-points lie on one positively semi-asymptotic and negatively semi-pervasive trajectory.

If we now introduce another equilibrium point A_0 on a different trajectory of (11), this gives for (13) another σ-point A_0 and two pervasive-asymptotic trajectories, one of each species. Hence $\mathcal{S}$ has now the σ-points everywhere dense and μ-trajectories of both species.

If instead of introducing A_0 in this manner we treat a complete trajectory (instead of a half-trajectory) of (11) in the previous manner, this is split into an infinite number of equilibrium points A_n and finite open arcs $A_n A_{n+1}$ $(n = \cdots -2, -1, 0, 1, 2, \cdots)$, where the A_n can be so chosen that the aggregate $\{A_n\}$ has A_0 as sole limit point. Choosing E so as to be positive on $\mathcal{S}$ except at the A_n, the stratum $\mathcal{S}$ is of type (c), with everywhere dense σ-points and no μ-points.

MELBOURNE UNIVERSITY,
AUSTRALIA.

REFERENCES.

E. Artin [1], "Ein Mechanisches System mit quasiergodischen Bahnen," *Abh. Math. Sem. Hamburg*, bd. 3 (1923), p. 170.

G. D. Birkhoff [1], "Quelques théorèmes sur le mouvement des systèmes dynamiques," *Bull. Soc. Math. France*, t. 40 (1912), p. 305.

[2], *Dynamical Systems*, New York (1927).

G. D. Birkhoff and P. A. Smith [1], "Structure analysis of surface transformations," *Jour. de Math. Pures et Appl.*, [9], t. 7 (1928), p. 345.

C. Carathéodory [1], "Über den Wiederkehrsatz von Poincaré," *Sitz. preuss. Akad. de Wiss.*, Berlin (1919), p. 580.

J. Hadamard [1], "Sur certaines propriétés des trajectoires en Dynamique," *Jour. de Math. Pures et Appl.*, [5], t. 3 (1897), p. 331.

H. Hilmy [1], "Sur les ensembles quasi-minimaux dans les systèmes dynamiques," *Annals of Math.*, vol. 37 (1936), p. 899.

E. Husson [1], *Les trajectoires de la dynamique*, Paris (1932).

H. Poincaré [1], "Mémoire sur les Courbes définies par une Équation différentielle," *Jour. de Math. Pures et Appl.*, [3], t. 8 (1882), p. 251.

[2], *Les Méthodes Nouvelles de la Mécanique Céleste*, t. 3 (Paris, 1899), Chap. 26.

THEORY OF QUASI-GROUPS.*

By B. A. Hausmann and Oystein Ore.

The present paper contains an investigation of the theory of groupoids, i. e. algebraic systems in which one operation is defined. There exists already a number of publications on such systems and various interesting although disconnected and scattered results have been obtained. In most of these studies certain axiomatic rules have been laid down and the properties of the corresponding systems have been deduced.

Here we shall try to obtain a more systematic theory by a different approach to the problem. Instead of starting from a chosen set of axioms we analyse the conditions necessary in order to obtain specific theories and theorems. In the end this turns out to be mainly an analysis of the associative law and the various possible formulations of the associative law represent the conditions for the existence of various analogues to the theorems in ordinary group theory.

In the first chapter one finds a classification of the groupoids and the definition of quasi-groups. Then follows an analysis of the conditions for co-set expansions with various properties, and one is led to three different types of associative laws. An interesting phenomenon by such quasi-groups is the existence of a minimal quasi-group not equal to a unit element but contained in all other quasi-groups.

In the second chapter one finds the structure theory of quasi-groups. We define normal quasi-groups and consider the conditions for laws of isomorphism. Next we determine the conditions for the normal quasi-groups to form a Dedekind structure and from these investigations follows from the general structure theory the conditions for the theorem of Jordan-Hölder and the direct decomposition theorem of Schmidt-Remak. Conditions for the existence of quotient groupoids are also given. Various problems which we had to omit from these considerations have been mentioned in the concluding section.

Chapter I. Quasi-Groups and the Associative Law.

1. Definitions and axioms. Since there exists a great variety of nomenclatures in our subject it may be well to specify at the beginning the definitions of the fundamental concepts used in the following.

* Received May 23, 1937.

We consider systems $\mathfrak{A}$ consisting of an arbitrary finite or infinite number of elements $a_1, a_2, \cdots$. Such a system shall be called a *groupoid* if it satisfies the

PRODUCT AXIOM. *To any two elements a' and a'' in $\mathfrak{A}$ there corresponds a unique third $a = a' \cdot a''$.*

Hence groupoids may be said shortly to be *systems with one operation.* The product axiom may be expressed also in the symbolic form

$$\mathfrak{A} \cdot \mathfrak{A} = \mathfrak{A}^2 \leqq \mathfrak{A}. \tag{1}$$

The *order* of a groupoid is the number of its elements. A subset $\mathfrak{B}$ of $\mathfrak{A}$ shall be called a *sub-groupoid* if

$$\mathfrak{B}^2 \leqq \mathfrak{B}. \tag{2}$$

Any set of elements of

$$a_1, a_2, \cdots$$

will generate a sub-groupoid

$$\mathfrak{A}' = \{a_1, a_2, \cdots\}$$

consisting of all possible products with a finite number of factors of the a_i. Since the associative law usually does not hold one has to distinguish the various ways of combining the factors. The simplest case is the *cyclic groupoid*

$$\mathfrak{C} = \{a\}$$

consisting of all possible products of one element. A *power* of a is any element belonging to $\{a\}$.

The *union* $[\mathfrak{B}, \mathfrak{C}]$ of two sub-groupoids $\mathfrak{B}$ and $\mathfrak{C}$ is the groupoid generated by the elements of $\mathfrak{B}$ and $\mathfrak{C}$ and the *cross-cut* $(\mathfrak{B}, \mathfrak{C})$ is the groupoid of their common elements. It is easily seen that the union and cross-cut as here defined satisfy the ordinary structure axioms. If we agree to consider the void set as a sub-groupoid if there exists sub-groupoids without common elements we have:

THEOREM 1. *The sub-groupoids of a given groupoid form a structure.*

Usually we shall have to impose further axiomatic restrictions on the groupoids considered. We shall say that $\mathfrak{A}$ is a *homogeneous groupoid* if it satisfies the axiom:

HOMOGENEITY AXIOM. *Each element of $\mathfrak{A}$ is a product.*

This means that for each a in $\mathfrak{A}$ we can determine other elements x and y such that

$$a = x \cdot y.$$

This again may be expressed as

(3) $$\mathfrak{A}^2 = \mathfrak{A}$$

which is a stronger form of (1).

When $\mathfrak{A}$ is a non-homogeneous groupoid it may be decomposed into disjoint classes by the following method. A product

(4) $$a = a_1 \cdot a_2 \cdots a_n$$

shall be said to have the *length* n. Let us again recall that since the associative law does not hold the product (4) is not uniquely defined but depends on the way in which the factors are combined. We define $\mathfrak{A}^n$ as the set of all elements of $\mathfrak{A}$ representable as products of length n. Then $\mathfrak{A}^n$ is seen to be a sub-groupoid and we have

$$\mathfrak{A} \geqq \mathfrak{A}^2 \geqq \mathfrak{A}^3 \geqq \cdots.$$

The elements of $\mathfrak{A}$ which are not products belong to

$$\Delta_1(\mathfrak{A}) = \mathfrak{A} - \mathfrak{A}^2$$

and in general those elements which are representable as products of length n but not of length $n+1$ belong to

$$\Delta_n(\mathfrak{A}) = \mathfrak{A}^n - \mathfrak{A}^{n+1}.$$

This gives us the decomposition of the elements of $\mathfrak{A}$ into the disjoint classes

$$\mathfrak{A} = \Delta_1(\mathfrak{A}) + \Delta_2(\mathfrak{A}) + \cdots.$$

Obviously those elements of $\mathfrak{A}$ which are representable as products of arbitrary length belong to all $\mathfrak{A}^n$. Hence we may say:

THEOREM 2. *The cross-cut*

$$\mathfrak{D} = (\mathfrak{A}, \mathfrak{A}^2, \cdots)$$

is the maximal homogeneous sub-groupoid of the given groupoid $\mathfrak{A}$.

It is clear that any other homogeneous sub-groupoid must be contained in $\mathfrak{D}$. If for some n we have $\mathfrak{A}^n = \mathfrak{A}^{n+1}$ then $\mathfrak{D} = \mathfrak{A}^n$ and we may call n the *order of non-homogeneity of* $\mathfrak{A}$.

If $\mathfrak{B}$ and $\mathfrak{C}$ are sub-groupoids of $\mathfrak{A}$, let us denote by $\mathfrak{B} \cdot \mathfrak{C}$ the totality of elements $b \cdot c$. Those sub-groupoids $\mathfrak{B}$ of $\mathfrak{A}$ for which

$$\mathfrak{A} \cdot \mathfrak{B} \leqq \mathfrak{B} \qquad (\mathfrak{B} \cdot \mathfrak{A} \leqq \mathfrak{B})$$

shall be called *left-ideal* (right-ideal) sub-groupoids. The study of these

groupoids would lead us to a multiplication theory of ideals and rings. In the present paper we are however more interested in systems approaching groups where such ideal groupoids do not exist.

The next considerations lead us to systems in which the following condition is fulfilled:

EXISTENCE OF QUOTIENTS. *To each pair a and b in $\mathfrak{A}$ there exist at least one x and one y such that*

$$a \cdot x = b, \qquad y \cdot a = b.$$

This condition may also be expressed

$$a \cdot \mathfrak{A} = \mathfrak{A} \cdot a = \mathfrak{A} \tag{5}$$

for every a. The solutions of the equations (5) are unique if we impose further:

CANCELLATION LAW. *When*

$$ax = ay \quad \text{or} \quad xa = ya \tag{6}$$

then $x = y$.

The two preceding conditions together may then be formulated as a single axiom:

QUOTIENT AXIOM. *To each pair a and b there exists a unique x and y such that*

$$ax = b, \qquad ya = b. \tag{7}$$

A groupoid satisfying the quotient axiom shall be called a *quasi-group.* Quasi-groups differ from the ordinary groups in the properties that no unit element, hence no inverse need exist and the associative law is usually not satisfied. For finite groupoids the cancellation law and the quotient axiom are equivalent since they both imply that each element of $\mathfrak{A}$ is contained in every column and line of the corresponding Cayley square. If one only supposes that one of the equations (7) are uniquely solvable we may call the corresponding system a left-hand or right-hand quasi-group.

2. Co-set expansions. Since a great part of the theory of groups depends upon the theory of co-set expansions it seems natural to investigate when a co-set expansion can be obtained in a groupoid. Let $\mathfrak{A}$ be the given groupoid and $\mathfrak{B}$ any sub-groupoid. The totality of elements $a \cdot \mathfrak{B}(\mathfrak{B} \cdot a)$ shall be called a left (right) *co-set* of $\mathfrak{A}$ with respect to $\mathfrak{B}$. We define:

Co-set expansion. *We shall say that there exists a left co-set expansion of $\mathfrak{A}$ with respect to $\mathfrak{B}$ when the co-sets of $\mathfrak{A}$ with respect to $\mathfrak{B}$ form a system of disjoint sets exhausting $\mathfrak{A}$. Furthermore no co-set shall contain equal elements.*

These conditions may be expressed

I. The equation

$$x \cdot b = a$$

has a solution for all a and some b in $\mathfrak{B}$.

II. Any relation

$$a_1 \cdot b_1 = a_2 \cdot b_2, \qquad b_1, b_2 \text{ in } \mathfrak{B}$$

implies

$$a_1\mathfrak{B} = a_2\mathfrak{B}.$$

III. Any relation

$$ab_1 = ab_2$$

implies $b_1 = b_2$.

Condition I implies namely that every element of $\mathfrak{A}$ belongs to some co-set while II shows that two co-sets are either identical or disjoint and III shows that no co-set contains equal elements.

From I follows immediately:

Theorem 3. *Co-set expansions can only exist in homogeneous groupoids.*

Condition III permits us to draw the ordinary conclusion that if $\mathfrak{A}$ and $\mathfrak{B}$ have the finite orders $N_{\mathfrak{A}}$ and $N_{\mathfrak{B}}$ then $N_{\mathfrak{B}}$ divides $N_{\mathfrak{A}}$ and the quotient $N_{\mathfrak{A}}/N_{\mathfrak{B}}$ is equal to the number of distinct co-sets.

Let us next determine the conditions for the existence of left co-set expansions for all sub-groupoids of $\mathfrak{A}$. When one agrees to consider $\mathfrak{A}$ itself as a sub-groupoid, it can be shown:

Theorem 4. *The necessary and sufficient condition that there shall exist a left co-set expansion for all sub-groupoids of a given groupoid $\mathfrak{A}$ is that:*

I. *To each a and b there shall exist such a b_a in the cyclic groupoid* $\{b\}$ that the equation

$$x \cdot b_a = a$$

has a solution.

II. *Any relation*

$$a_1 \cdot b_1 = a_2 \cdot b_2$$

shall imply

$$a_1\mathfrak{B} = a_2\mathfrak{B}$$

where $\mathfrak{B}$ is any groupoid containing b_1 and b_2.

III. *The left-hand cancellation law shall be satisfied.*

The proof is immediate. Let us observe that the first condition implies that every element in $\mathfrak{A}$ lies in a co-set with respect to any cyclic sub-groupoid of $\mathfrak{A}$ and conversely when it is satisfied any a lies in some co-set with respect to any sub-groupoid.

In the case where one does not consider $\mathfrak{A}$ to be a sub-groupoid the above conditions must be slightly modified.

3. The first associative law. From now on we shall suppose that $\mathfrak{A}$ is a *quasi-group.* Some of the results are however valid under somewhat more general conditions but we shall leave these more general formulations of the theorems to the reader.

We shall also suppose that $\mathfrak{A}$ is of finite order. This assumption is not essential in the sense that one can formulate the results in such a manner that they also hold in the infinite case. The theory is considerably simplified however for finite quasi-groups and the number of necessary conditions is reduced. One consequence of this condition is that every sub-groupoid $\mathfrak{H}$ is a quasi-group. We have namely for any h in $\mathfrak{H}$

$$h\mathfrak{H} = \mathfrak{H}h = \mathfrak{H}$$

because all these complexes contain the same number of elements.

From theorem 4 follows next:

THEOREM 5. *The necessary and sufficient condition that there exist left co-set expansions in a finite quasi-group $\mathfrak{A}$ with respect to any sub-groupoid is that the following condition be fulfilled:*

A. CONDITION OF ASSOCIATIVITY. *Any relation*

$$a_1 \cdot b_1 = a_2 \cdot b_2 \tag{8}$$

shall imply

$$a_1 \cdot \mathfrak{B} = a_2 \cdot \mathfrak{B} \tag{9}$$

where $\mathfrak{B}$ is any groupoid containing b_1 and b_2.

We shall now analyse the condition of associativity and show that it may actually be considered as a weak form of the associative law.

THEOREM 6. *The condition of associativity in a quasi-group* $\mathfrak{A}$ *is equivalent to:*

A_1. ASSOCIATIVE LAW. *Let* a *and* b *be arbitrary elements of* $\mathfrak{A}$. *For the fixed element* c_0 *let* d_0 *be determined such that*

$$(a \cdot b) \cdot c_0 = a \cdot d_0. \tag{10}$$

Then we have for any c

$$(a \cdot b) \cdot c = a \cdot d \tag{11}$$

where d *belongs to the groupoid* $\mathfrak{C} = \{c_0, d_0, c\}$ *generated by* c_0, d_0 *and* c.

Proof. Obviously the condition A implies A_1. To prove the converse let a relation (8) hold. Since we can write $a_1 = a_2 \cdot x$ it can be expressed

$$(a_2 \cdot x) \cdot b_1 = a_2 \cdot b_2$$

and hence by A_1

$$(a_2 \cdot x) \cdot c = a_1 \cdot c = a_2 \cdot d$$

where d belongs to the groupoid $\{b_1, b_2, c\}$. Hence we have for any groupoid $\mathfrak{B}$ containing b_1 and b_2

$$a_1 \cdot \mathfrak{B} \leqq a_2 \cdot \mathfrak{B}$$

and on account of the symmetry we obtain (9).

It may be of interest to observe that more special associative laws are implied in (11).

THEOREM 7. *The associative law* A_1 *implies*

$$(a \cdot b) \cdot c = a \cdot d$$

where d *belongs to the groupoid* $\mathfrak{C} = \{e_{ab}, b, c\}$ *generated by* b, c *and the element* e_{ab} *defined by*

$$(ab) e_{ab} = ab.$$

The proof of this theorem follows by identifying the last relation with (10).

4. The second associative law. The associative law A_1 gave the necessary and sufficient condition for the existence of left co-sets in a quasi-group. By imposing further conditions on the properties of the co-sets we are led to more special associative laws.

THEOREM 8. *The necessary and sufficient conditions that the co-sets shall have the property that any element in the co-set defines the same co-set is that the quasi-group have the following associative law:*

A_2. *For any set of elements a and b one has*

$$(a \cdot b) \cdot c = a \cdot d \tag{12}$$

where d belongs to $\{b, c\}$.

Proof. In order that

$$a \cdot \mathfrak{B} = (a \cdot b)\mathfrak{B} \tag{13}$$

for any groupoid $\mathfrak{B}$ containing b the condition (12) must obviously be satisfied and when (12) is satisfied (13) must also hold. The associative law A_2 implies the condition of associativity A (or A_1). If namely (8) holds we find from (13) for any $\mathfrak{B}$ containing b_1 and b_2

$$a_1 \cdot \mathfrak{B} = (a_1 \cdot b_1)\mathfrak{B} = (a_2 \cdot b_2)\mathfrak{B} = a_2 \cdot \mathfrak{B}$$

and (9) is proved.

When the associative law A_2 holds one can deduce several other important properties of the quasi-group.

THEOREM 9. *For any sub-groupoid* $\mathfrak{B}$ *containing b one has*

$$a \cdot \mathfrak{B} = (a \cdot b) \cdot \mathfrak{B} = (a \cdot \mathfrak{B}) \cdot b. \tag{14}$$

These complexes are all contained in $a \cdot \mathfrak{B}$ and contain the same number of elements. Furthermore:

THEOREM 10. *Any co-set* $a \cdot \mathfrak{B}$ *remains the same when multiplied by an element of* $\mathfrak{B}$ *on the right. The co-set contains its multiplier a.*

To prove the last property we observe that according to (14) we can determine an e_a in $\mathfrak{B}$ such that

$$a \cdot b = (a \cdot e_a) \cdot b$$

and when b is cancelled one finds

$$a = a \cdot e_a. \tag{15}$$

This proof gives another interesting property of the quasi-groups with the associative law A_2. In any quasi-group one can find a unique element e_a satisfying (15). We shall call e_a the *right unit* of a. The totality of right units generate a sub-groupoid which we shall call the *right unit groupoid* $\mathfrak{E}$. From (14) follows that in the case where A_2 holds the unit e_a is contained in any sub-groupoid of $\mathfrak{A}$.

THEOREM 11. *In quasi-groups satisfying the associative law A_2 the right-unit groupoid is a minimal groupoid contained in all other sub-groupoids.*

The existence of such a minimal groupoid not equal to a unit element is somewhat surprising. Let us illustrate its existence by the example of a quasi-group defined by the following multiplication table:

	1	2	3	4	5	6
1	2	1	3	5	4	6
2	1	3	2	4	6	5
3	3	2	1	6	5	4
4	5	4	6	2	1	3
5	4	6	5	1	3	2
6	6	5	4	3	2	1

This groupoid is cyclic. Its minimal groupoid consists of the elements 1, 2, 3.

THEOREM 12. *In a quasi-group satisfying the associative law A_2 let a and b be given elements. Then there exists a power b_1 of b such that*

$$(a \cdot b) \cdot b_1 = a. \tag{16}$$

The proof of this theorem follows from theorem 9 and 10 when applied to $\mathfrak{B} = \{b\}$. The relation (16) represents a weak form for *inverses.*

Among the other properties of the right unit elements in such systems let us mention

$$(a \cdot b) \cdot e_c = a \cdot b_1, \qquad (a \cdot e_b) \cdot c = a \cdot c_1$$

where b_1 and c_1 are powers of b and c respectively. If a quasi-group satisfying A_2 contains an *idempotent* element

$$a^2 = a$$

then a is an *absolute right unit* since the minimal groupoid consists of a single element a.

5. The third associative law. There exists other properties of the ordinary co-sets in groups which we cannot derive from A_2. In a group we have: Let $\mathfrak{B} > \mathfrak{C}$ be two subgroups of a given group $\mathfrak{A}$. The co-set decomposition of $\mathfrak{A}$ with respect to $\mathfrak{C}$ may be obtained by first decomposing $\mathfrak{A}$ with respect to $\mathfrak{B}$ and then substituting in the co-sets $a \cdot \mathfrak{B}$ the co-sets of $\mathfrak{B}$ with respect to

$\mathfrak{C}$. This property may be called the *transitivity* of the co-set decomposition. If we want the co-set expansions in quasi-groups to have the same property we must have

$$a \cdot (b \cdot \mathfrak{C}) = a_1 \cdot \mathfrak{C} \tag{17}$$

where a_1 is a fixed element independent of the particular element chosen in $\mathfrak{C}$. Since we want the transitivity in addition to the former properties of the co-set decomposition we can assume that A_2 holds. When $\mathfrak{C}$ is taken as the right unit groupoid $\mathfrak{E}$ we find from (17)

$$a(be_b) = ab = (ab)e_{ab} = a_1 \cdot e$$

where e is some element of $\mathfrak{E}$. Since $\mathfrak{E}$ is contained in all sub-groupoids we find from condition A

$$a(b \cdot \mathfrak{C}) = (ab)\mathfrak{C}.$$

This shows, when $\mathfrak{C}$ is taken as a cyclic group, that *the necessary and sufficient condition that the co-set decomposition in a quasi-group $\mathfrak{A}$ be transitive and have the property that every element of a co-set define the same co-set, is that the following associative law be satisfied:*

A_3. Associative law. *For any three elements a, b, c*

$$a(bc) = (ab)c_1$$

where c_1 is a power of c.

In connection with the various associative laws which we have derived in the preceding one should mention the generalisations of the associative law considered by other authors. Most of these investigations differ from the present since they deal with associative laws in rings, hence with associative laws containing two operations. One of the most interesting purely multiplicative associative laws is due to *Suschkewitsch:*[1]

In the relation

$$(a \cdot b) \cdot c = a \cdot d \tag{18}$$

the element d does not depend on a.

This law is a special case of the second associative law A_2 as one sees by making a an element of the groupoid $\{b, c\}$. In this case there also exists an absolute right unit. To see this we write

$$(a \cdot e_a) \cdot c = a \cdot c$$

[1] A. Suschkewitsch, "On a generalization of the associative law," *Transactions of the American Mathematical Society*, vol. 31 (1929), pp. 204-214.

and according to (18) we have for any b

$$(b \cdot e_a) \cdot c = b \cdot c$$

or

$$b \cdot e_a = b.$$

This shows that $e = e_a = e_b = \cdots$ is a right unit. On the basis of the associative law (18) Suschkewitsch proves the theorem of *Lagrange* for quasi-groups, i. e. the order of a sub-quasi-group divides the order of the quasi-group. This is of course true whenever co-set decompositions exist. Suschkewitsch also considers the *associative law*

$$(a \cdot b) \cdot c = a \cdot (b \cdot c_1) \tag{19}$$

where c_1 *only depends on* c. This is a special case of our associative law A_3. The most interesting result in his paper is that his quasi-groups may be obtained by certain substitutions from ordinary groups.

Another system of associative laws was considered by Miss Moufang.[2] She considers quasi-groups in which there exists a two-sided absolute unit e, for every element a there exists an inverse a^{-1}

$$a^{-1} \cdot a = a \cdot a^{-1} = e$$

with the further property that

$$a \cdot (a^{-1} \cdot b) = b, \qquad (b \cdot a^{-1}) \cdot a = b.$$

In addition one shall have the associative laws

$$(a \cdot (c \cdot a)) \cdot b = a \cdot (c \cdot (a \cdot b)) \tag{20}$$

$$(a \cdot b) \cdot (c \cdot a) = a \cdot ((b \cdot c) \cdot a). \tag{21}$$

It is then shown that when the associative law (20) holds any *two* elements a and b generate an ordinary group. If both (20) and (21) hold, then any *three* elements for which $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ will also generate a group.

Another method of studying the properties of quasi-groups is to consider those elements a which satisfy one of the relations

$$(x \cdot y) \cdot a = x \cdot (y \cdot a)$$
$$(x \cdot a) \cdot y = x \cdot (a \cdot y)$$
$$(a \cdot x) \cdot y = a \cdot (x \cdot y)$$

for all x and y. Such investigations have been made by *Schönhardt*.[3]

---

[2] Ruth Moufang, "Struktur von Alternativkörpern," *Mathematische Annalen*, vol. 110 (1935), pp. 416-430.

[3] E. Schönhardt, "Über Lateinische Quadrate und Unionen," *Journal für Mathematik*, vol. 163 (1930), pp. 183-230.

Chapter II. Normal Quasi-Groups.

1. Definitions and associative law. Since the structural properties of ordinary groups are closely connected with the properties of normal sub-groups, we shall consider the possibility of defining normal sub-groupoids of a quasi-group. It is natural to define the *transform* $b^{(a)}$ of an element b with respect to a as the solution of the equation

$$a \cdot b = b^{(a)} \cdot a.$$

A *normal sub-groupoid* $\mathfrak{N}$ of $\mathfrak{A}$ shall then be defined as a groupoid containing the transforms of all its elements. This is equivalent to

$$a \cdot \mathfrak{N} = \mathfrak{N} \cdot a \tag{1}$$

or

$$a \cdot n = n' \cdot a \tag{2}$$

for all n and n' in $\mathfrak{N}$ and a arbitrary in $\mathfrak{A}$.

THEOREM 1. *The cross-cut* $(\mathfrak{M}, \mathfrak{N})$ *of two normal sub-groupoids is normal or vacuous.*

For an element d in the cross-cut one has

$$a \cdot d = n \cdot a = m \cdot a$$

and hence $n = m = d'$ is also in the cross-cut. Through a similar argument one finds:

THEOREM 2. *If* $\mathfrak{N}$ *is normal and* $\mathfrak{B}$ *any sub-groupoid then the cross-cut* $(\mathfrak{B}, \mathfrak{N})$ *is vacuous or normal in* $\mathfrak{B}$.

To obtain further properties of the normal sub-groupoids it is necessary to introduce some form of the associative law. We shall begin by assuming only:

N_1. ASSOCIATIVE LAW. *For any four elements we have*

$$(ab)(cd) = (\bar{a}b_1)\bar{d} = \bar{\bar{a}}(b_2\bar{\bar{d}}) \tag{3}$$

where $\bar{a}$ *and* $\bar{\bar{a}}$ *are powers of* a, $\bar{d}$ *and* $\bar{\bar{d}}$ *powers of* d *and* b_1 *and* b_2 *belong to the groupoid* $\{b, c\}$.

By making d an element of $\{b, c\}$ we find as a special case of N_1

$$(ab)c = \bar{a} \cdot b_1 \tag{4}$$

where $\bar{a}$ is a power of a and b_1 belongs to $\{b, c\}$. Similarly one finds

$$a(bc) = a_1 \cdot \bar{c} \tag{5}$$

where $\bar{c}$ is a power of c and a_1 belongs to $\{a, b\}$.

These laws are weaker than A_2 and they are not sufficient to obtain co-set expansions. They are however already sufficient to prove several facts about normal sub-groupoids.

We shall say that two sub-groupoids $\mathfrak{B}$ and $\mathfrak{C}$ are *permutable* if one always has for their elements

$$b \cdot c = c' \cdot b'.$$

Hence a normal sub-groupoid is permutable with every other groupoid.

THEOREM 3. *Let $\mathfrak{B}$ and $\mathfrak{C}$ be two permutable groupoids. Then the elements of their union*

$$\mathfrak{U} = [\mathfrak{B}, \mathfrak{C}] \tag{6}$$

have one of the forms

$$u = b, \quad u = c, \quad u = b \cdot c. \tag{7}$$

Proof. The union $\mathfrak{U}$ consists of all elments of $\mathfrak{B}$ and $\mathfrak{C}$ and elements obtained by repeated products of these. Hence the theorem may be deduced by induction from the following relations which are all consequences of the associative laws (3), (4) and (5):

$$\begin{aligned} b_1 \cdot (b \cdot c) &= b_2 \cdot \bar{c} \\ c_1 \cdot (b \cdot c) &= c_1 \cdot (c_2 \cdot b_2) = c_3 \cdot \bar{b}_2 \\ (b_1 c_1)(bc) &= (c_2 b_2)(bc) = \bar{c}_2(b_3 \bar{c}) = \bar{c}_2(c_4 b_4) = c_5 b_5. \end{aligned}$$

As a corollary it follows from theorem 3 that if $\mathfrak{B}$ is normal and $\mathfrak{C}$ arbitrary the union (6) will consist of elements of the form (7).

A fundamental result is the following:

THEOREM 4. *Let $\mathfrak{B}$ and $\mathfrak{C}$ be permutable groupoids and $\mathfrak{D}$ some groupoid containing $\mathfrak{B}$. Then the three groupoids satisfy the Dedekind relation*

$$(\mathfrak{D}, [\mathfrak{B}, \mathfrak{C}]) = [\mathfrak{B}, (\mathfrak{D}, \mathfrak{C})]. \tag{8}$$

Proof. Obviously the right-hand side of (8) is contained in the left.

To show the converse we observe that each element of $[\mathfrak{B}, \mathfrak{C}]$ has one of the forms (7). Hence the left-hand cross-cut consists of elements satisfying one of the relations

$$d = b, \; d = c, \; d = bc.$$

In the two first cases the element is contained in $\mathfrak{B}$ or $(\mathfrak{D}, \mathfrak{C})$. In the last case b belongs to $\mathfrak{D}$ since $\mathfrak{D} > \mathfrak{B}$, and since $\mathfrak{D}$ is a quasi-group the element c must belong to $\mathfrak{D}$, hence to $(\mathfrak{D}, \mathfrak{C})$. Finally (8) also holds when one of the cross-cuts is void.

2. The law of isomorphism. Let $\mathfrak{B} > \mathfrak{C}$ be two sub-groupoids of $\mathfrak{A}$. With each such pair we associate a *quotient structure* consisting of all sub-groupoids $\bar{\mathfrak{C}}$ of $\mathfrak{A}$ such that

$$\mathfrak{B} \geqq \bar{\mathfrak{C}} \geqq \mathfrak{C}.$$

Furthermore we shall say that two quotients are *structure isomorphic*

$$\mathfrak{B}/\mathfrak{C} \sim \mathfrak{B}_1/\mathfrak{C}_1$$

when there exists a one-to-one correspondence between their sub-groupoids $\bar{\mathfrak{C}} \rightleftarrows \bar{\mathfrak{C}}_1$ preserving union and cross-cut. We can then prove:

THEOREM 5. *Let $\mathfrak{B}$ and $\mathfrak{C}$ be groupoids such that $\mathfrak{C}$ is normal in the union $[\mathfrak{B}, \mathfrak{C}] = \mathfrak{U}$ and the cross-cut $(\mathfrak{B}, \mathfrak{C}) = \mathfrak{D}$ is not void. Then $\mathfrak{D}$ is normal in $\mathfrak{B}$ and there exists a structure isomorphism*

$$[\mathfrak{B}, \mathfrak{C}]/\mathfrak{C} \sim \mathfrak{B}/(\mathfrak{B}, \mathfrak{C}). \tag{9}$$

Proof. From theorem 2 follows that $\mathfrak{D}$ is normal in $\mathfrak{B}$. Next let $\bar{\mathfrak{D}}$ and $\bar{\mathfrak{C}}$ be any groupoids such that

$$\mathfrak{B} \geqq \bar{\mathfrak{D}} \geqq \mathfrak{D}, \qquad [\mathfrak{B}, \mathfrak{C}] \geqq \bar{\mathfrak{C}} \geqq \mathfrak{C}. \tag{10}$$

We shall then show that the correspondence

$$\bar{\mathfrak{D}} \to [\mathfrak{C}, \bar{\mathfrak{D}}], \qquad \bar{\mathfrak{C}} \to (\mathfrak{B}, \bar{\mathfrak{C}}) \tag{11}$$

defines the structure isomorphism (9). One finds first from the Dedekind relation (8) that

$$\bar{\mathfrak{D}} = (\mathfrak{B}, [\mathfrak{C}, \bar{\mathfrak{D}}]), \qquad \bar{\mathfrak{C}} = [\mathfrak{C}, (\mathfrak{B}, \bar{\mathfrak{C}})]$$

so that (11) gives a one-to-one correspondence between the two structures. From these relations one obtains

$$[\bar{\mathfrak{D}}_1, \bar{\mathfrak{D}}_2] \to [\mathfrak{C}, \bar{\mathfrak{D}}_1, \bar{\mathfrak{D}}_2] = [\bar{\mathfrak{C}}_1, \bar{\mathfrak{C}}_2]$$
$$(\bar{\mathfrak{D}}_1, \bar{\mathfrak{D}}_2) \to [\mathfrak{C}, (\bar{\mathfrak{D}}_1, \bar{\mathfrak{D}}_2)] = [\mathfrak{C}, (\mathfrak{B}, [\bar{\mathfrak{C}}_1, \bar{\mathfrak{C}}_2])] = [\bar{\mathfrak{C}}_1, \bar{\mathfrak{C}}_2]$$

completing the proof.

Some other properties of these normal sub-groupoids may be derived. One finds that their properties correspond exactly to those of the semi-normal elements introduced by *Ore* [4] for an arbitrary structure. One can derive the same weak theorems for chains of normal sub-groupoids as for chains of semi-normal elements in structures. These properties are however not strong enough to derive a theorem of Jordan-Hölder, even in the weak form that two normal chains have the same length.

From now we shall consider quasi-groups with both right and left co-set expansions. Hence we shall assume as in chap. I:

A_2. Associative law. *For any three elements*

$$(ab)c = a \cdot d, \qquad a(bc) = f \cdot c$$

where d belongs to $\{b, c\}$ and f to $\{a, b\}$.

This assumption has various consequences. We have seen that it implies the existence of a minimal sub-groupoid $\mathfrak{C}$ contained in all other sub-groupoids and $\mathfrak{C}$ is generated by all right and left units of the elements of $\mathfrak{A}$. In this case the cross-cut of two groupoids cannot be void. Theorem 3 takes the simpler form: *If $\mathfrak{B}$ and $\mathfrak{C}$ are permutable then the union $[\mathfrak{B}, \mathfrak{C}]$ consists* of the elements $b \cdot c$.

By making $b = e_a$ in (3) one finds that N_1 now implies:

A_4. Associative law. *For any three elements*

$$(ab)c = \bar{a}(\bar{b}\bar{c})$$

where $\bar{a}$, $\bar{b}$, $\bar{c}$ are powers of a, b, c respectively.

In order to combine A_2 and A_4 we are led to:

A_5. *For any three elements a relation*

$$(ab)c = a(\bar{b}\bar{c}) \tag{12}$$

shall hold both ways, where $\bar{b}$ is a power of b and conversely, $\bar{c}$ a power of c and conversely. Similarly we assume

[4] Oystein Ore, "On the theorem of Jordan-Hölder," *Transactions of the American Mathematical Society*, vol. 41 (1937), pp. 266-275.

$$a_1(b_1c_1) = (\bar{a}_1 \cdot \bar{b}_1) \cdot c_1. \tag{13}$$

This law shall be assumed to hold in the following. Obviously it implies A_2 and A_4. It also implies N_1 as one sees from the relations

$$(ab)(cd) = ((ab)\bar{c}) \cdot \bar{d} = (a(\bar{b}\bar{c})) \cdot \bar{d}.$$

Let us now return to theorem 5. We expand $\mathfrak{B}$ in co-sets with regard to $\mathfrak{D}$

$$\mathfrak{B} = \mathfrak{D} + b_2\mathfrak{D} + \cdots + b_k\mathfrak{D}. \tag{14}$$

We can then show that

$$\mathfrak{U} = \mathfrak{C} + b_2\mathfrak{C} + \cdots + b_k\mathfrak{C} \tag{15}$$

is a co-set expansion of $\mathfrak{U}$. First no two co-sets (15) can be equal because it would lead to the equality of two co-sets (14). Secondly every element $b \cdot c$ of $\mathfrak{U}$ belongs to a co-set, because if one writes $b = b_i \cdot d$ one finds

$$b \cdot c = (b_i \cdot d)c = b_i \cdot c_1.$$

Thus we have defined a one-to-one correspondence

$$b_i\mathfrak{D} \rightleftarrows b_i\mathfrak{C} \tag{16}$$

between the co-sets of the two quotients (9). This correspondence is seen to define the structure isomorphism (11) between the two quotients because if

$$\bar{\mathfrak{D}} = \mathfrak{D} + \bar{b}_2\mathfrak{D} + \cdots + \bar{b}_t\mathfrak{D}$$

then one finds that the corresponding $\bar{\mathfrak{C}}$ in (11) is given by

$$\bar{\mathfrak{C}} = \mathfrak{C} + \bar{b}_2\mathfrak{C} + \cdots + \bar{b}_t\mathfrak{C}.$$

This shows furthermore that the indices of corresponding quotients $\bar{\mathfrak{D}}/\mathfrak{D}$ and $\bar{\mathfrak{C}}/\mathfrak{C}$ are the same. Such a correspondence of structures may be called a *strong* structure isomorphism $(\mathfrak{A} \simeq \mathfrak{A}')$.

THEOREM 6. *Let* $\mathfrak{C}$ *be normal in* $[\mathfrak{B}, \mathfrak{C}]$. *Then* $(\mathfrak{B}, \mathfrak{C})$ *is normal in* $\mathfrak{B}$ *and there exists a strong structure isomorphism*

$$[\mathfrak{B}, \mathfrak{C}]/\mathfrak{C} \simeq \mathfrak{B}/(\mathfrak{B}, \mathfrak{C})$$

defined by a one-to-one correspondence between the co-sets of the two quotients.

3. Decomposition theorems. We shall continue to assume the associative law A_5. We prove first:

where $\bar{a}$ is a power of a and conversely. These quasi-normal groupoids correspond exactly to the quasi-normal groups introduced by *Ore*[6] in ordinary groups and a similar theory may be developed. For a large part of the theory it is sufficient to assume the associative law A_4. Let us observe that this law from the point of view of the theory of structures has similar properties to those of the ordinary associative law, since by the change of parentheses the elements are left in the same groupoid.

One can also develop the theory of two-sided co-sets in quasi-groups. If $\mathfrak{H}$ and $\mathfrak{K}$ are two sub-groupoids, then the totality of elements in the complex

$$\mathfrak{H}(a\mathfrak{K}) \tag{24}$$

is called a *two-sided co-set.* It is obvious that every element of a quasi-group belongs to some two-sided co-set (24). We shall say that $\mathfrak{A}$ possesses a *two-sided co-set expansion* with respect to $\mathfrak{H}$ and $\mathfrak{K}$ if

$$\mathfrak{A} = \mathfrak{H}(a_1\mathfrak{K}) + \mathfrak{H}(a_2\mathfrak{K}) + \cdots$$

where any co-sets

$$\mathfrak{H}(a\mathfrak{K}), \qquad \mathfrak{H}(b\mathfrak{K})$$

defined by arbitrary a and b are either identical or have no element in common.

One can now derive necessary and sufficient conditions for the existence of two-sided co-set expansions, with given properties much in the same manner as we have done in chap. I for one-sided co-sets. We shall not specify the various associative laws which are necessary for the various properties. Most of them are associative laws on four elements. One is also led to:

A_C: Associative law. *For any three elements*

$$a(bc) \doteq (\bar{a}b) \cdot \bar{c}$$

where $\bar{a}$ and $\bar{c}$ are powers of a and c and conversely.

This law does not imply one-sided co-set expansions. If one wishes both one- and two-sided co-set expansions one is again led to the associative law A_3.

Theorem 16. *Let us suppose that the associative law A_3 holds. Then the quasi-group $\mathfrak{A}$ has two-sided co-set expansions with respect to any two sub-groupoids $\mathfrak{H}$ and $\mathfrak{K}$. Furthermore*

$$\mathfrak{H}(a\mathfrak{K}) = (\mathfrak{H}a)\mathfrak{K}.$$

---

[6] Oystein Ore, "Structures and group theory," *Duke Mathematical Journal*, vol. 3 (1937), pp. 149-174.

Any element in a two-sided co-set defines the same co-set and a co-set remains the same by left multiplication with an element of $\mathfrak{H}$ *or right multiplication with an element of* $\mathfrak{K}$.

Another important problem is the study of the special theory of *Abelian quasi-groups.* Equally interesting is also the question of the conditions for the validity of the *Sylow theory of groupoids.* All these problems must however be reserved for a later occasion.

Other problems of interest are the determination of all quasi-groups with given associative laws and their relation to ordinary groups. By certain operations one can derive quasi-groups from ordinary groups. A special case of such quasi-groups is given by the quasi-groups considered by *Suschkewitsch.*

YALE UNIVERSITY.

CORRECTIONS.

OTTO SZÁSZ, *On the partial sums of certain Fourier series.*

P. 705, l. 3, the formula on this line should be numbered (31);

P. 705, l. 6 from bottom, read $[K_q(t-x) - K_q(t+x)]$ instead of $[K_q(t-x) + K_q(t+x)]$;

P. 706, l. 1, read $\sum\limits_{0<\lambda_n\leq\mu}$ instead of $\sum\limits_{0<\lambda_n\leq\omega}$.